Inclusive Education Evaluation by using Big Data Sources of Information (descriptive analytics)

Inception Report

UNICEF Kazakhstan

Oxford Policy Management

28 March 2023

Table of Contents

1. About this inception report

1. As per the initial Request for Proposals (RfP) for this study, UNICEF's "Inclusive Education Evaluation by using Big Data Sources" had two objectives:
   a) **Understand the perspectives of different stakeholders on the inclusive education landscape in Kazakhstan by social media listening tools** to identify barriers, assess what might be working (not working) and propose recommendations for promoting best practices.
   b) Use spatial data in Kazakhstan to disaggregate regional variation in inclusive education outcomes across different levels of schooling.
2. During the contracting phase, **UNICEF and OPM together decided to restrict the scope of the study to cover the first objective (1 a) above.**

1. About this inception report

1. As part of the inception phase, OPM and UNICEF held multiple discussions on how this innovative evaluation could **complement an ongoing formative evaluation commissioned by UNICEF** that seeks to assess how state education programmes have contributed to inclusive education over the past decade.

2. Additionally, both parties took this opportunity to **finalize the definition of inclusive education** that this study will focus on, **finalize the evaluation questions** (EQs), and deliberate over the **proposed methodology** and the infrastructure support that would be required to execute this innovative evaluation.

3. **This Inception Report is a contracted deliverable and presents a synthesis of the outcomes of these discussions**. Note that during contract negotiations, OPM and UNICEF together decided to replace a long-text format inception report with a slide deck.

4. **OPM's background and capabilities in implementing** projects similar to the one at hand is outlined in Annex A.

The **remainder of this report is structured as follows:** we first discuss how we **define and operationalise** inclusive education for this study (**section 2**), before providing some context on inclusive education and the use of social media in Kazakhstan in **section 3**. We then discuss evaluation objectives and questions in **section 4**. **Section 5 and 6** present our methodology of how we intend to answer these questions. We discuss caveats and limitations in **section 7**, before turning to ethical considerations (**section 8**), and our implementation timeline in **section 9**.

2. Defining ‘Inclusive Education’ for this study

1. **To answer our evaluation questions** (presented in section 4)**, we first need to define and operationalize “inclusive education” in the context of this study.** We start by looking at the potential target group that ‘inclusive education’ relates to.
2. **Potential target group:** Children with special educational needs in Kazakhstan’s context include (see slides 10 ff.) :
   a) Children with behavioral and emotional problems, unfavorable psychological factors;
   b) Children with barriers of socio-psychological, economic, linguistic and cultural nature;
   c) Children with developmental disabilities.
3. **Our focus: Our work will focus on category c) above** and considers a) and b) if they are connected to the third condition. A key rationale of focussing on one of these categories is that it allows us to **conduct focussed and well-defined social media listening analyses**, as explained further below.

2. Defining 'Inclusive Education' for this study

3. **Categorizing disabilities from a public policy viewpoint: We consider both the *medical model* as well as the *social model* of disability in our conceptualisation of disability for social media listening.** The former (medical model) is based on the characteristics of the disease and disorders. The social model acknowledges that disability is a multi-dimensional, complex interaction of health and contextual factors that enable or restrict children from participating effectively (see Box 1).

Box 1. An example of the medical versus social model of disability

Aigul is a secondary school student whose eyesight is deteriorating because of diabetic retinopathy, a common complication of diabetes. The medical model sends her to a clinic but there is no treatment to halt the progressive damage to her sight. It becomes more difficult for her to get to school using public transport and her parents are trying to find extra money to afford taxis. Aigul is excluded because she cannot read the texts or afford the extra transport costs. She is very worried that she will have to drop out of school.

The social model finds out what assistance she needs in order to continue to attend school and keep up with her studies. This means finding solutions to overcome the barriers to her inclusion. For example, her teachers and the Department of Education could lobby the IT developers to make scanner apps for blind and visually impaired users available in her language. At the same time the occupational therapist can work with Aigul, her teachers and classmates to find and practice with her the most practical route to school by public transport. Aigul can also apply for a disability allowance to cover the additional disability-related costs.

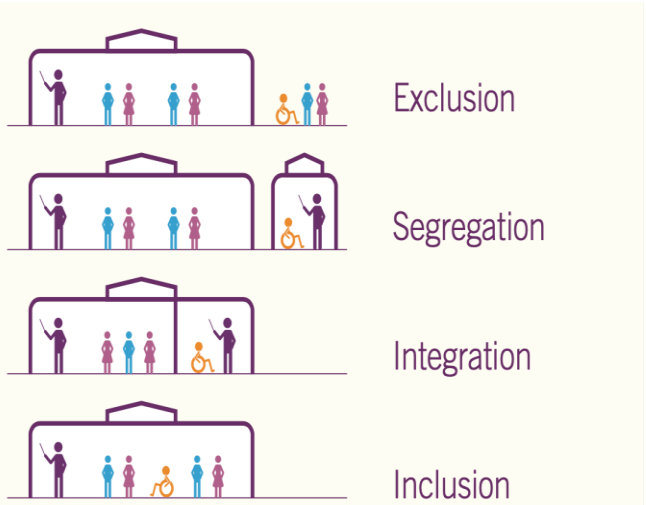
Aigul has an impairment (deteriorating eyesight) which limits her functioning (reading textbooks and using public transport) until the participation restriction is lifted (by having access to the state-of-the-art scanner apps and access to support to learn new skills for independent travel).

2. Defining 'Inclusive Education' for this study

4. **Operationalizing 'inclusion': Inclusion means more than being physically present at school** and there are important distinctions between exclusion, segregation, integration and inclusion (Figure 1).

Exclusion: no recognition of the right or capacity of children with disabilities to education and consequent denial of access to education in any form.
Segregation: placement of children with disabilities in separate environments designed or used to respond to particular or various impairments, in isolation from children without disabilities.
Integration: placement of children with disabilities in existing mainstream educational institutions, as long as the child can adjust to fit the standardized requirements of such institutions.
Inclusion: a process of systemic reform embodying changes and modifications in content, teaching methods, approaches, structures and strategies in education to overcome barriers. It requires a commitment to changing the system to fit the student.

Figure 1. Conceptual difference between exclusion, segregation, integration and inclusion



Source: UNESCO; UNICEF; Global Partnership for Education; (UK) Foreign, Commonwealth & Development Office, (2021) Education Sector Analysis Methodological Guidelines Volume III. Page 35

2. Defining 'Inclusive Education' for this study

5. **Human Rights view of inclusive education :** Article 24 of the *Convention on the Rights of Persons with Disabilities (CRPD)* focuses on the right to an inclusive education system *"States parties must ensure* ***the realization of the right of persons with disabilities to education through an inclusive education system at all levels****, including preschool, primary, secondary and tertiary education, vocational training and lifelong learning, extracurricular and social activities, and for all students, including persons with disabilities, without discrimination and on an equal basis with others."*

→ **For the purpose of this study, we will restrict our analysis to children in the first three levels of preschool, primary, secondary education.**

6. **Key international conventions and declarations on disability and inclusive education:** List available in Annex B (see Box 2). **These frameworks will inform our analysis of social media data and our interpretation of results.**

2. Defining ‘Inclusive Education’ for this study: Summary

- In summary, for this study, our online search will focus on social media conversations that tackle the following issue:

- For children with other forms of barriers to accessing education, such as behavioral and emotional issues or those of a socio-psychological, economic, linguistic and cultural nature, these conditions will be covered if they are related to developmental disabilities and access to pre-school, primary, secondary education.

3. Context: Inclusive education landscape in Kazakhstan

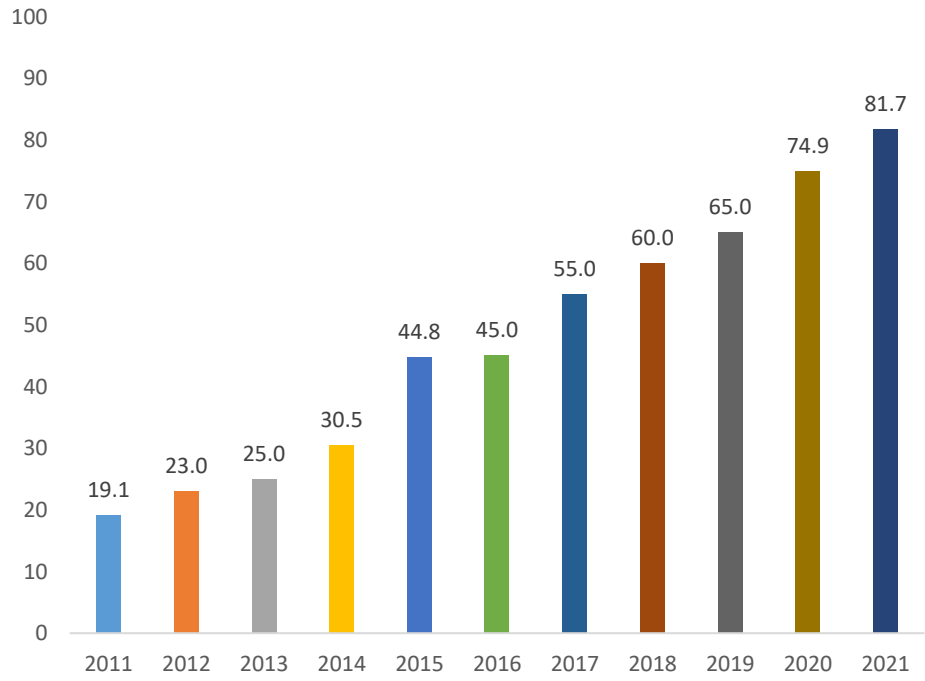
1. From the information shared by UNICEF Kazakhstan, we know the following about the inclusive education landscape in the country:

- Kazakhstan is a relatively young country with close to **34% population below 18 years of age**. Currently it has over **160,000 children with *special needs*** and this number is growing every year.
- With this increase, **the network of special educational organizations that provide educational services and support has also been rising**. For example, in the 2020-2021 academic year, 403 special education organizations were active in the country.
- For the 2020-2021 academic year, close to **39% of disabled children were reportedly attending regular schools or attending classes online**, a figure 12 percentage point higher than these levels in 2016-2017.

3. Context: Inclusive education landscape in Kazakhstan

- In 2021, **about 82% of schools were reported to have created suitable conditions to facilitate inclusive education**, up from 19% in the academic year 2011 (Figure-2).
- In 2020, to facilitate inclusivity, **teachers-assistants** were introduced at all levels of education to provide pedagogical support for children with disabilities. According to the Ministry of Education's website, there were about 1,500 teacher-assistants who worked at school in late 2022 (September 8th, 2022).

Figure-2 : Percentage of schools creating suitable conditions for inclusive education



Source: Ministry of Education, National report on the state and development of the education system KZ, 2023

3. Context: Inclusive education landscape in Kazakhstan

2. a. General barriers to inclusive education

- According to the **Ministry of Education, Kazakhstan (2020), major barriers** that impede inclusive education and learning quality includes factors like **lack of acknowledgement of the needs of students and teachers**; **dearth of qualified school specialists**; **overcrowded classrooms**; **lack of redressal of issues like bullying in schools**; and **low levels of interactions between teachers and parents**.
- Prominent **barriers specific** to **inclusive education** include **lack of appropriate schooling infrastructure** (like ramps, handrails etc.); **difficulties for parents in accessing relevant information** for children with disabilities; **lack of appropriate psychological environment** and **support** at schools that can address issues of stigma and discrimination; **shortage of transport facilities**, especially in the context of large distances to secondary schools and **lack of employment opportunities** for the disabled, among others (Ministry of Education, Kazakhstan (2020).

3. Context: Inclusive education landscape in Kazakhstan

2. **b. Gender specific barriers to inclusive education**

- For instance, about 33% of schools in Kazakhstan have **outdoor toilets**, which are often **not disabled friendly**, suffer from sanitation problems, and are associated with the **additional risk of higher sexual violence against girls** (CABAR, 2019).
- **Higher gender-based violence** further **reduced mobility of girls with disabilities** thereby serving as a major barrier to inclusive education (United Nations Population Fund (UNFPA), 2020).
- These **shortcomings are further accentuated by social norms**, which consider **household chores** to be the **'duty of women'** as a result of which girls in general and especially those with disabilities are more likely to not enroll of dropout from school (UNFPA, 2020).

3. Context: Inclusive education landscape in Kazakhstan

3. In terms of the legal framework, there is no separate law on inclusive education in Kazakhstan but there exists a President's Law on amendments and additions to some legislative acts of the Republic of Kazakhstan that relates to inclusive education (here).

- As per the earlier version of Law on Education (2007), children with special educational needs were defined as ***"persons (children) who experience permanent or temporary difficulties in obtaining education due to health, who need special, general education programs of additional education"***. Thus, this definition of children with special educational needs led to understanding that inclusive education was interpreted and identified only with respect to **children with disabilities**.
- However, in 2021 the definition of children with special educational needs was changed: "*persons (children) with special educational needs – persons (children) who have permanent or temporary needs in special conditions for obtaining an appropriate level of education and additional education*". This corresponds to a broader understanding of inclusive education.

3. Context: Inclusive education landscape in Kazakhstan

- **This 2021 law has become an important document for providing children with special needs with equal access to quality education.** The law regulates the activities of psychological and pedagogical support services for children with special educational needs in general education schools, flexibility of curricula, programs, as well as the system for evaluating student achievements depending on the characteristics of their development.
- **However, there is no concrete description regarding categories of children who can be defined as children with special educational needs in the Law of Education.** According to the Rules of psychological and pedagogical support in educational organizations (Decree 6 of Ministry of Education (here)), there are three categories of children with special needs that are as follows:

3. Context: Inclusive education landscape in Kazakhstan

i. **Children with behavioral and emotional problems, unfavorable psychological factors** (violations of upbringing in the family, child-parent and intra-family relations);

ii. **Children with barriers of socio-psychological, economic, linguistic and cultural nature** (pedagogical neglect of children from families of social risk, children experiencing difficulties of adaptation in society (families of refugees, migrants, repatriates);

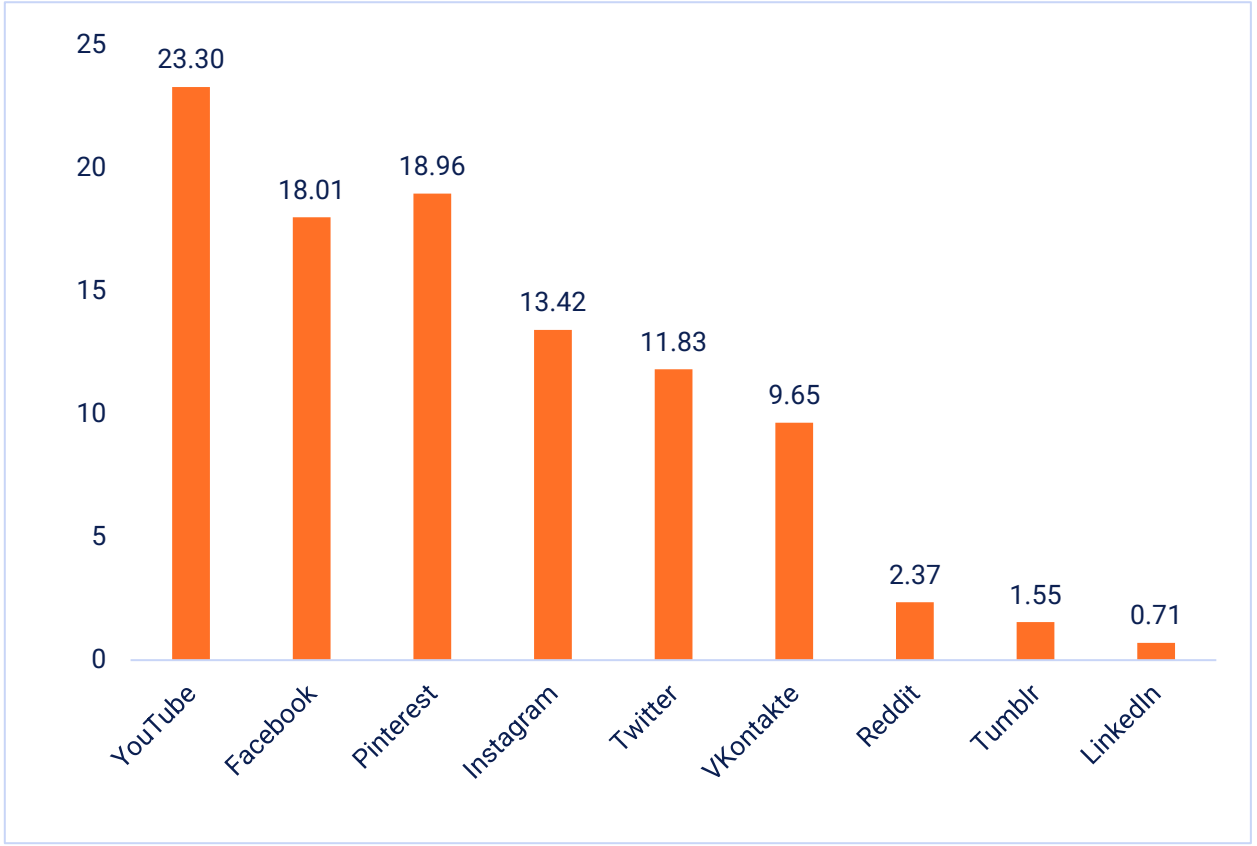
iii. **Children with developmental disabilities (hearing, vision, intelligence, speech, musculoskeletal system, mental retardation and emotional-volitional disorders).**

→ Our work focusses on category iii. above but considers i. and ii., if they are connected to the third condition.

3. Context: Social media use in Kazakhstan

- As of January 31, 2022, there were about **13.8 million estimated social media users in Kazakhstan**
- This figure corresponds to about **72% of the overall population**, although note that all these users are not necessarily unique
- **Youtube, Facebook, Pinterest** and Instagram were the most popular platforms over the past year (see Figure 3)
- Analyzing data from these platforms has a vast potential in informing diverse opinions and voices on vital socio-economic issues

Figure 3: Monthly average market share from January 22 to January 23



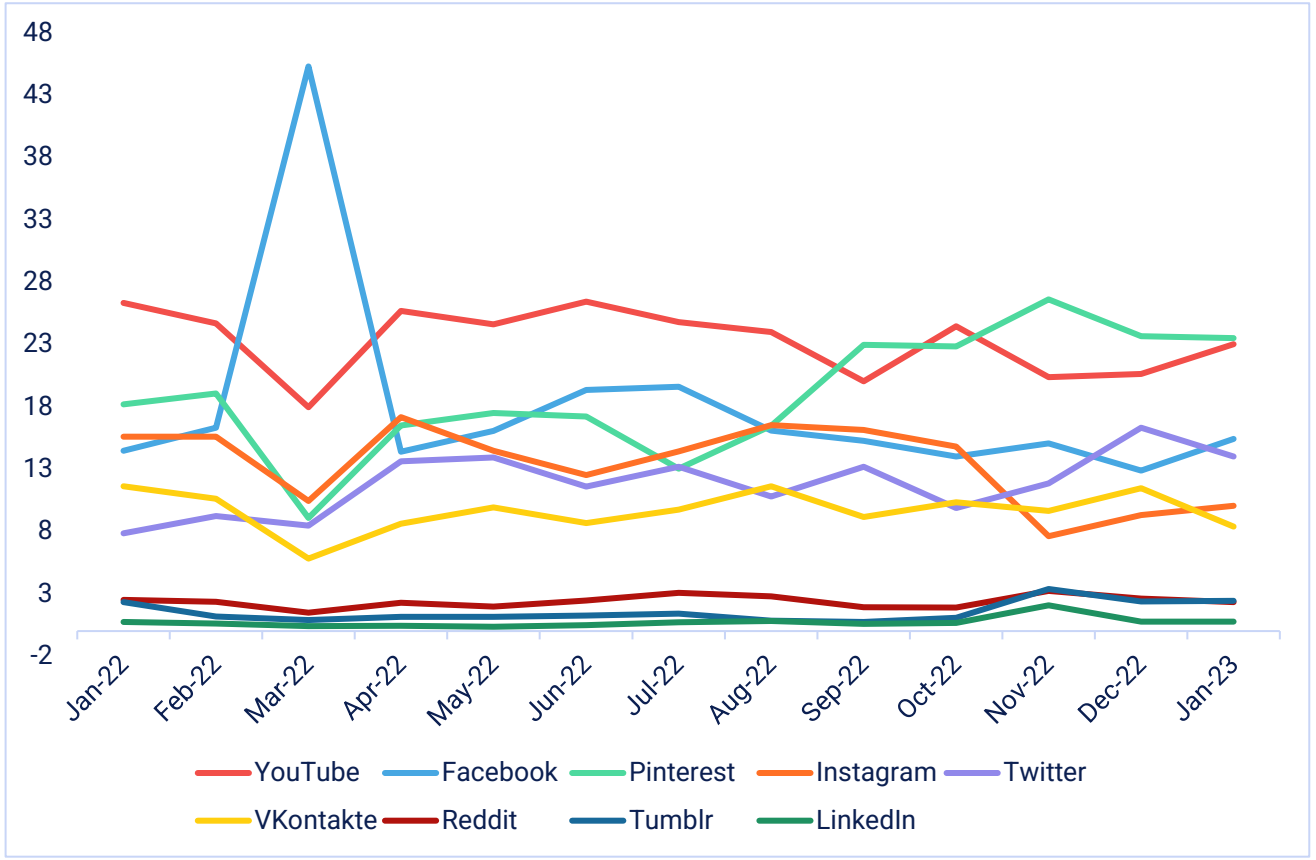
Source: https://datareportal.com/reports/digital-2022-kazakhstan

3. Context: Social media use in Kazakhstan

- **Youtube (23%), Vkontakte's (10%)** market share has been mostly stable over the past year (Figure-4)
- **Pinterest (19%) and Twitter's (12%)** market share has been rising steadily, while Instagram (13%) has conceded some market share in the past months (Figure-4)
- Literature shows that close to **99 % of those between 18 and 29 years use social media** (Sairambay, 2022)
- About 3 in 4 young people in this age group use social media for domestic and international news (Sairambay, 2022)

→ **Social Media Listening will aim to capture data from main platforms. It is also likely that we will be able to access posts produced by youth, which will give a unique perspective to our analysis.**

Figure 4: Social Media Market Share in Kazakhstan (January 22 to January 23)



Source: https://gs.statcounter.com/social-media-stats/all/kazakhstan

4. Evaluation purpose, objectives, and questions: Purpose

1. **Our primary purpose will be to describe the perspectives of individuals, in particular parents, teachers, and the youth overall, on the status of inclusive education in Kazakhstan**, using online (social) media listening tools.
2. Our hypothesis is that by analysing this non-traditional data, we will be able to **gain insights and formulate recommendations based on viewpoints from different stakeholders that are otherwise difficult to observe or are neglected when studied through traditional methods and approaches.**
3. A **formative evaluation** to assess the contribution of different state education programmes in promoting inclusive education is currently underway (implemented by *Junction Bulgaria (JB)*).

4. Evaluation purpose, objectives, and questions : Purpose

4. Rather than looking at efficacy of specific programmes, **our study will complement this work** and provide evidence on the inclusive education landscape in Kazakhstan by using information from publicly accessible social media sources.
5. In this context, **this study does not have a conventional Evaluation Matrix or a Theory of Change (ToC).** This is because we do not explicitly test the validity of a ToC with this study. If of interest, a reference to the draft ToC produced by JB is provided in Annex G.
6. However, the study design ensures that the **proposed evaluation questions** (discussed subsequently) **are mapped to the objectives of the study, as mentioned in the Terms of Reference (ToR)**, via rigorous methodological tools that are consistent with the relevant literature on social media listening.

4. Evaluation purpose, objectives, and questions : Purpose

7. **Our study will also contribute to the emerging evidence on the use of innovative analytical methods, such as text analytics, in evaluations.** In Kazakhstan, earlier, studies have, for example, studied connectivity disparity across schools using big data (see here, Office of Global Innovation, UNICEF Kazakhstan).

8. **Emerging evidence can then be used by the Government of Kazakhstan** to address existing gaps and inform the next education policy framework and contribute to the implementation of the national project "Educated Nation" 2021-2025.

9. This is in tandem with the **SDG 4 goal of "*Ensuring inclusive and equitable quality education for all*"**. Our ambition is that our evidence will inform the following:
   - discourse on inclusive education;
   - identify barriers;
   - policy changes that are working and those that are needed to achieve inclusive education outcomes.

4. Evaluation purpose, objectives, and questions : Objectives

1. As described, the **primary objective** of our study will be to **describe and summarize** the perspectives of individuals, especially parents, teachers, and youth, on the status of and barriers to inclusive education in Kazakhstan, using online social media listening tools.

2. The **secondary objectives** of this research will be to:

- Summarize **recent trends in conversations on inclusive education, including an analysis of barriers to accessing education for children with disabilities being mentioned.**
- Where possible, implement **comparative analyses to bring out any contrast in discussions across different stakeholders**, especially from groups that may otherwise have a comparatively smaller digital footprint.
- Where possible, **investigate whether state programmes** are being mentioned in the context of online discussions on inclusive education.

3. Consistent with the UN Convention on the Rights of the Child (CRC), particularly Article 28, and the UN Convention on the Rights of Persons with Disabilities, particularly Articles 24, 29 and 30, **we adopt a human rights-based approach to this study** whereby the focus of our research is on equal opportunities for disabled children in accessing pre-primary, primary and secondary education institutions.

4. Evaluation purpose, objectives, and questions: Evaluation Questions

1. The specific **Evaluation Questions** (EQs), consistent with the objectives previously discussed are as follows:

a) EQ1) A 'description' of online discussions on inclusive education in Kazakhstan: **What is the content of discussions on inclusive education in Kazakhstan? How does this vary by the 'stakeholder group' being looked at?** (*See methods section: we suggest analysing social media data from two stakeholder groups: the general public and stakeholders active in inclusive education in Kazakhstan.*) **How does it vary over time?**

b) EQ2) An analysis of what 'barriers' or 'gaps' in inclusive education are being mentioned**: What do people mention as factors 'holding back' inclusive education?** How does this vary by 'stakeholder group' being looked at?

c) EQ3) **Do people mention state programs at all?** How does this vary by the 'stakeholder group'?

4. Evaluation purpose, objectives, and questions: What this study does not cover

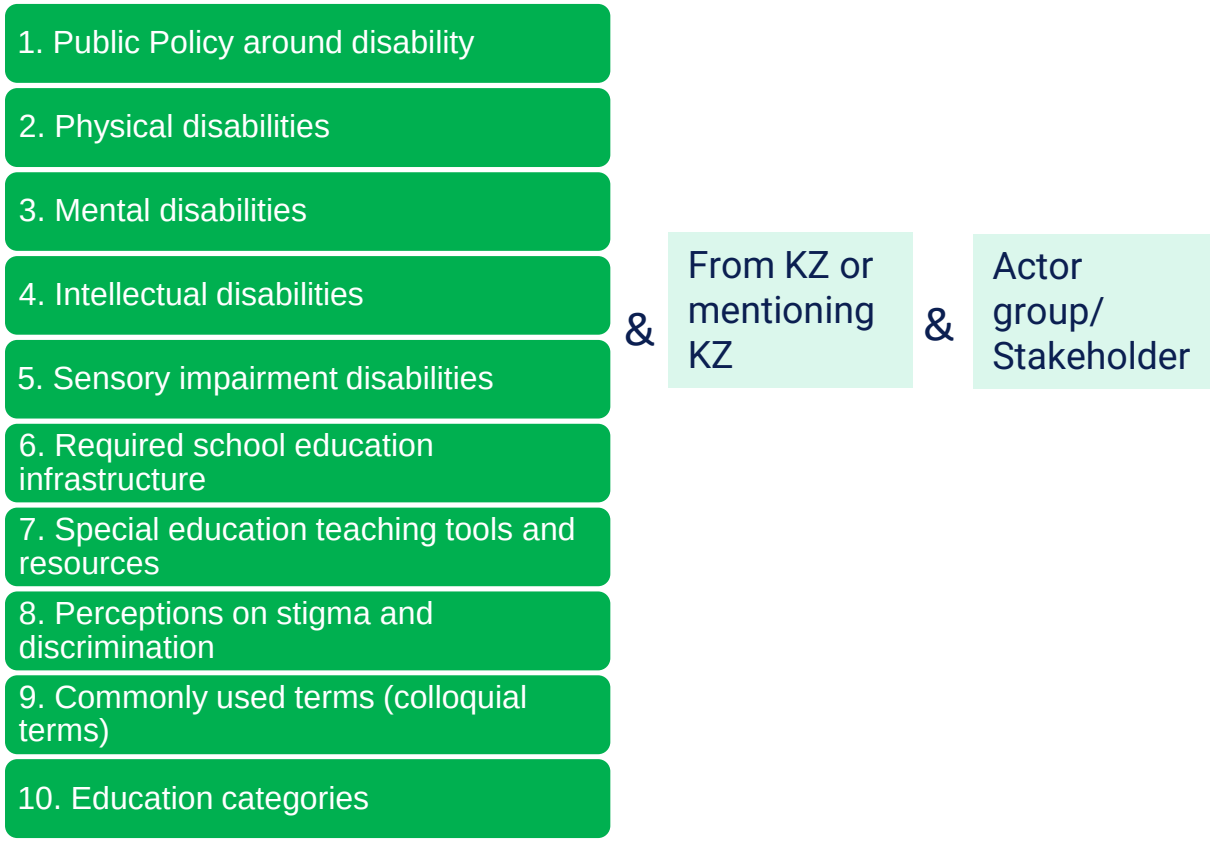
1. As discussed in the previous sections, this study does not cover the following:

- **An evaluative assessment of state programmes directly.** This work is already being undertaken via the formative evaluation being conducted by the Junction Bulgaria team.
- **A disaggregation of analyses across geographies**. Our experience with social media data shows that the geo-locations are often masked by users due to a variety of considerations like privacy among others. As a result, social media data can often not be disaggregated meaningfully to provide heterogeneity analysis by geographic units.
- The study is restricted in scope to cover disability induced barriers to inclusive education at pre-primary, primary and secondary levels. This is done mainly to focus on inclusivity in education and have a well-defined, targeted approach to designing the search queries (discussed subsequently).

5. Methodology: Tentative domains of search queries

- **To scrape social media data from the Internet, we need to 'translate' the previously identified definition into search queries.** This means that we need to find a list of words that together describe our topic of interest.
- **As discussed, we focus on publicly accessible online text on the topic of 'inclusive education',** i.e. text where different stakeholders talk about access to pre-primary, primary, and secondary education for children with disabilities.
- To make the task of finding a list of words that describe this manageable, **we suggest breaking the topic of 'inclusive education' into sub-topics** (See Figure 5 for illustration purpose).
- **Each sub-topic will consist of a long query of words in English, Kazakh and Russian**. This will be manually validated by colleagues in KZ.
- Our main search would combine the sub-topics with the appropriate Boolean operators (AND; OR) into one or multiple topics, that together will comprise our search for inclusive education discussions online.
- An example of query building is provided later in the Methodology section.

Figure 5. Illustration of how queries for data scraping will be formulated



5. Methodology: Two data scraping strategies

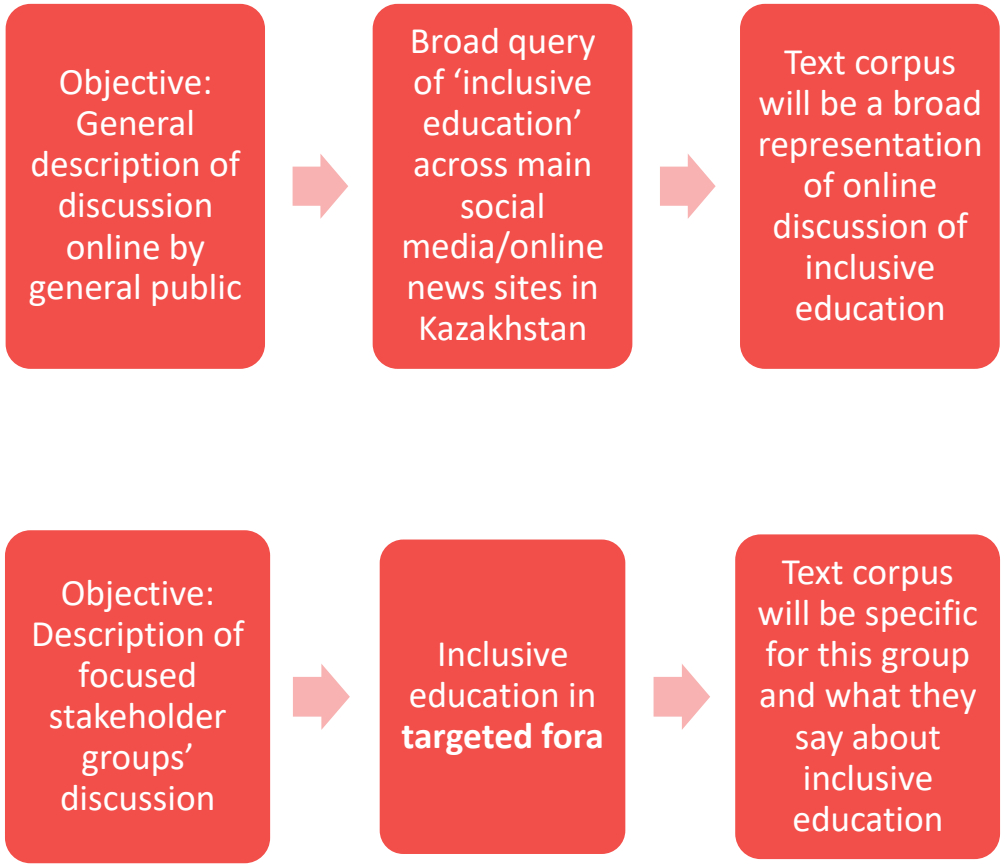
1. **The first step in any social media analysis is to build a text corpus to be analysed, by 'scraping' text from the Internet.** This involves searching for specific terms being mentioned in text – see slide 22 for how we conceptualise this for 'inclusive education' – and then downloading this text. We will provide an example for search queries on the next slide.

2. **As discussed previously, and in order to answer our EQs appropriately, we will choose two approaches to build text corpora for analysis (Figure 6).** These are summarized in Figure 3 and listed below. Both corpora will be screened and validated by OPM to verify if we are capturing the discussions online that we intend to study. If not, appropriate changes will be made to data scraping strategy.

i. **General public:** We pool scraped data from accessible social media platforms and analyse these sources together. Suitable when similar volume and quality of data available from all stakeholders.

ii. **Stakeholder group active in inclusive education:** The second approach will build a separate text corpus for a different stakeholder group. This is useful when trying to analyse groups with vastly different social media presence.

Figure 6. Choice of two analytical workflows



5. Methodology: Building search queries (example)

For instance (Figure-7), consider an example of a query that is built to cover:

- Physical disabilities **AND** specific languages (Kazakh, Russian, English) **AND** Country (Kz)

OR

- Different education categories of interest **AND** specific languages (Kazakh, Russian, English) **AND** Country (Kz)

***Figure 7. Illustration** of how queries for data scraping will be formulated*

1. Physical disabilities

- List of terms with “or” operators that describe or are used to refer to physical disabilities in Kazakh, Russian, English.

& Language & Country

2. Education categories

- List of terms that describe or are used to refer to different education categories in Kazakh, Russian, English (with ‘or’ operators).

5. Methodology: Building search queries (example) using Boolean operators for different languages

- Below, we provide three concrete **examples** for how search queries might look like. Note that these are illustrative to clarify how queries are built and should not be seen as finalized.
- The final list of **queries will cover our list of sub-topics**.

English

- (("school" or "learning" or "education") and ("disability" or "special needs" or "learning difficulties")) or ("inclusive education" or "disability inclusion") AND language:"en" AND country:"kz"

Kazakh

- (((мектеп OR оқыту OR білім) AND (мүгедек OR "ерекше қажеттіліктер" OR "оқудағы қиындықтар")) OR ("инклюзивті білім беру" OR ("мүгедектерді интеграциялау" AND жасөспірім))) AND language:"kz"

Russian

- (((школ* OR обучен* OR образован*) AND (инвалид* OR "особые потребности" OR "трудности в обучении")) OR ("инклюзивное образование" OR ("интеграция инвалидов" AND несовершеннолетн*))) AND language:"ru" AND country:"kz"

5. Methodology: Building search queries

Some important factors to keep in mind with respect to the queries are as follows:

1. **The finalized list of publicly accessible data sources, which includes social media or online news platforms will be shared with UNICEF for feedback.** The finalized list of topics and keywords will go through manual validation which is discussed subsequently. These results will also be shared with UNICEF for feedback and finalization of data that will then be used for analysis.
2. **This search will not be a one-time process but will need a few iterations to ensure that the relevant words, phrases, organizations and contexts are being covered across the relevant languages**, locations and other socio-economic and cultural factors that have a presence on social media. However, it is not ideal to go back and forth with this process so it will be undertaken only a few number of times.

5. Methodology: Validating search results

A key step in social media listening analysis is human validation of results obtained from running online search queries. The objective is to ensure that the text corpus does indeed contain text relating to the relevant search topic. In our case this is 'inclusive education in Kazakhstan' as previously defined.

- In human validation, **we read through a sample of results obtained from search queries** to see the kind of social media conversations being picked up in the searches.
- **Some of the keywords in our search queries may be picking up topics that are not relevant to the study (inclusion error)** and in some cases, important publicly available conversations on inclusive education may not be picked up (**exclusion error**) (Figure 8). Through this validation stage, we will attempt to minimize both.
- We will sample results per sub-topic (see slide 25) and let a human review these results. Note, however, because 'topics' and 'issues' that people talk about are inherently vague, we will always have some degree of inclusion or exclusion errors.
- **Inclusive and exclusion error will likely vary by the two corpora we will build.** Our search of 'general public' might suffer from more inclusion error, as it will capture a broad conversation online. A more 'limited' search per topic and limiting actors will possibly suffer from exclusion error. By looking at results from both scraping strategies, we get a comprehensive, and more robust, picture of online discussions on the general topic of inclusive education.

Figure 8. Contingency Table for the True vs Our Search

		Truth	
		Is not about topic	Is about topic
Our search	Is not about topic → exclude	Correct	Exclusion error
	Is about topic → include	Inclusion error	Correct

5. Methodology: SML infrastructure support

As mentioned in our Technical Proposal, we will be grateful to UNICEF for the support on access to Talkwalker for social media listening. TalkWalker is a SML platform from which search queries can be deployed and to which UNICEF has access. There are multiple advantages of using TalkWalker for this project, such as previous utilization of TalkWalker in evaluation work for UNICEF. During the inception phase, it became clear, however, that it is possible that the evaluation team will not be able get access to TalkWalker for this specific study.

Alternatively, we have therefore explored the possibility of the team using Meltwater. At the time of writing this report, we are confident that we will get access to Meltwater, an alternative platform that will also allow us to deploy SML search queries. Access to this platform will be possible via OPM.

- **TalkWalker/Meltwater are both popular platforms** that can help analysing online social media and online news data to generate insights on individual and institutional perspectives of inclusive education in Kazakhstan.
- **We now plan to use Meltwater to build our text corpus, download it, and then analyse is using our own software**.
- **Initial exploration of search queries during the inception phase, using Meltwater, has indicated that social media data from Kazakhstan can be accessed.** However, it might be possible that we will have to deploy other – additional – scraping strategies to gain access to some social media platforms in Kazakhstan. We will only identify the full extent of this requirement once we start exploring initial text corpora produced using Meltwater.

5. Methodology: Analytical infrastructure support

As mentioned in our Technical Proposal, we will be grateful to UNICEF for the support on access to the Azure Databricks server. There are multiple advantages of these for the project.

- **Databricks provides a collaborative work environment** that facilitates working on the same code, run such code in an environment that is shared by all team members and makes it efficient to disseminate results on data collection and data analysis.
- **Data collection tasks are typically time intensive, and it is impractical to run such scripts on local computers, which are subject to failures**. Databricks, instead, is particularly well suited to run scripts for data collection on a scheduled basis. Data analysis with NLP models can also be very CPU intensive and Databricks is specifically designed to parallelize such workloads. This may reduce considerably the time needed to execute the scripts.
- **Databricks allows to securely share access credentials to other resources**, like TalkWalker (if accessible), so that access keys never leave UNICEF's premises. Databricks assets that will be required includes a fully managed data science workspace, notebook environment, model management and a data lake.

6. Methodology: Analysis

- **The analysis will seek to answer the evaluation questions (EQ) that have been posed previously (see slide 20).** By analysis, we mean the analytical approaches that will be used on text data after obtaining the two corpora from implementing SML queries on TalkWalker or Meltwater.
- **Note that our EQ are mostly descriptive in nature.** Analytical methods used will follow from this descriptive objective.
- **Our analysis will be mainly based on Natural Language Processing (NLP) approaches**. The term NLP can be used synonymously with text analytics. A variety of different approaches exist. The applicability of some, more complex, approaches (sentiment analysis and topic modelling) will depend on the language of the text data to be analysed. Some are well adapted to operate well in English, but possibly less so for Russian and Kazakh.
- Likely analytical approaches to be used will be:
    1. Simple **comparisons of engagement indicators over time** (likes, volume, such as number of posts, mentions, etc.) for different stakeholder groups. Comparisons of such indicators across stakeholder groups. Where possible, comparisons across geography.

.

6. Methodology: Analysis

2. Simple **word counts and word clustering techniques** to describe the content of the text corpora, which will be calculated using Python libraries like Pandas and Spacy. Comparisons of these over time and across stakeholders. A focus here will be on words associated with state programmes, in order to tackle EQ3.
3. If possible, **topic modelling** approaches to identify themes being mentioned in the different text corpora. Comparisons of these over time and across stakeholders.
4. Where possible, **sentiment analysis** approaches to identify positive, negative, or neutral text related to 'inclusive education'. Comparisons of these over time and across stakeholders.

- The final choice of analytical approaches used will depend on an evaluation of the text corpus that we can build from our online search queries.
- Analysis will be conducted in two key ways:
  - First, we will use TalkWalker's or Meltwater's own analytical tools to describe **engagement indicators mentioned above.**
  - Second, other analyses will be implemented by using statistical programming tools: **Python and R.** All code will be made available to UNICEF after completion of this study.

7. Caveats and Limitation

It is also important to acknowledge the limitations of the methods used and flagging any potential risks associated with the same. Also, it is vital to reiterate the unique features of this innovative evaluation and how it is different from conventional studies that this study seeks to complement.

a) **This study will complement the work being undertaken by the JB team through their formative evaluation of different state education programmes in promoting inclusive education.** Instead of focusing on specific govt. programmes, we will provide perspectives of different stakeholders on inclusive education landscape and conduct sentiment analysis by using publicly accessible social media sources.

b) **As these perspectives are unknown a-priori and come from publicly accessible social media data that is different from data collected via traditional survey-based methods**, we will work with EQs stated previously but will stay away from the discussion on effect sizes that emerge from specific hypotheses central to conventional quantitative evaluation studies.

7. Caveats and Limitation

c) **For all the benefits that social media listening tools have to offer, our ability to analyze the publicly available information on social media is constrained by socio–economic abilities as well as the conscious choice of individuals and institutions to engage with social media platforms**. Thus, the findings emerging from our study, although helpful in understanding multifarious perspectives on inclusive education, are not representative at the country or any sub-group level.

d) For reasons discussed earlier, **the scope of the study is restricted to cover children with disabilities for levels of education covering pre-primary, primary and secondary education in Kazakhstan.**

e) **Our ability to analyse the social media data across geographies will depend on whether the data being scaped online is geo-tagged**. As many users online hide or mask their locations, these aspects may be beyond the scope of the study.

7. Caveats and Limitation

f) **As most social media platforms keep digital records only for a limited period, the analysis is likely to be restricted to one or two years in the recent past.**

g) We acknowledge that barriers faced by different groups in accessing inclusive education may differ systematically. For example, we are aware that gender is an important consideration in this regard. **However, our ability to meaningfully disaggregate analyses by different groups of interest using social media data is constrained by the difficulty in identifying the group or individual identity of the account posting social media posts (for example, if a tweet is posted by a man or woman) and the difficulty in identifying whether a post talks specifically about gender barriers in one way or another (for example, if a view is posted is general or about women or men in accessing inclusive education).** This is the reason that no disaggregation as part of the EQs or indicators is committed to in the inception phase. Depending on the extent of such identification, we will try to analyse the results by gender.

8. Ethical considerations

1. As a values-driven organization, OPM is always respectful to the rights of the participants of its research projects and has a policy to ensure complete adherence to research ethics.

2. **OPM follows a set of ethical principles in conducting all fieldwork that we have based on our own experience as per the United Nations Evaluation Group (UNEG) evaluation policy.** This study will adhere to international best practice standards in evaluation, including the OECD DAC. We do not plan to conduct any primary data collection as part of this project, but publicly available online data will be scraped from the Internet. Ethical considerations for using this online data will be taken into consideration, with a particular focus on **privacy protection.** This also means that we will anonymize identities of individuals as well as organizations.

8. Ethical considerations

3. As part of OPM's internal protocols, we considered the core principles outlined by the Government Social Research on use of social media data (2016).

4. As per our discussions with UNICEF, it was rechecked with the Guidance and Evaluation Specialist from ECA regional office that an ERB clearance would not be required for this project as we are using only publicly accessible social media and news data as well as anonymizing any personal level identifying information.

9. Deliverables, Timelines, Quality Assurance and Dissemination Plan

1. **The overall project deliverables has three vital components of which this inception slide deck is the first.**

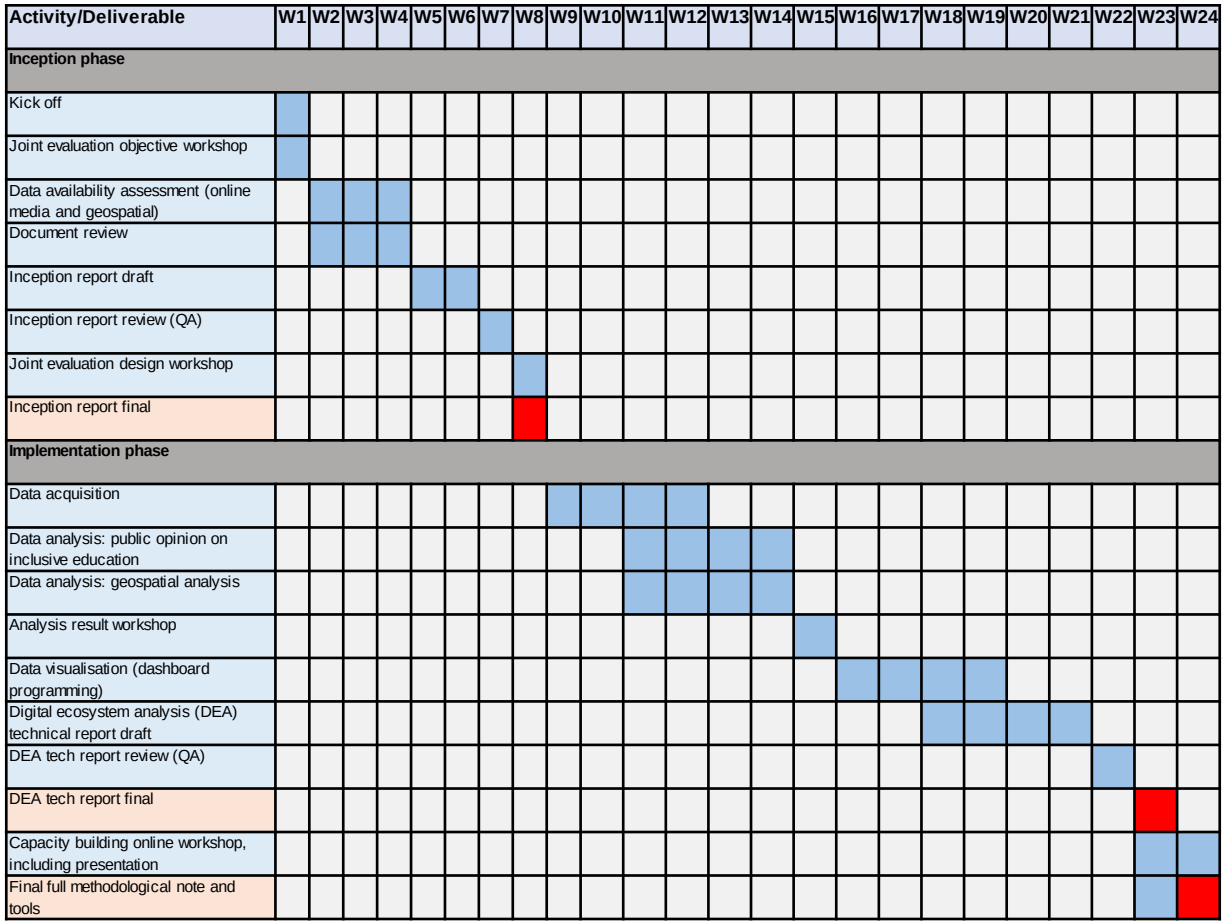
i. Inception slide deck- Due 8 weeks of the project timeline (week ending 11th February 2023).

ii. The Digital Ecosystem Analysis report – which we consider to be the main technical report that includes results from all key analyses. Due in 23 weeks of the project timeline (week ending 25th June 2023).

iii. The final 'package', including the final methodological note and tools. This also includes a capacity building workshop and the delivery of all analytical code. Due in 24 weeks of the project timeline (week ending 2nd July 2023).

9. Deliverables, Timelines, Quality Assurance and Dissemination Plan

Figure 9. Timelines



9. Deliverables, Timelines, Quality Assurance, and Dissemination Plan

1. OPM is committed to disseminating research findings from this study in a way that it can serve all the stakeholders associated with inclusive education in Kazakhstan.
2. We will implement a **three-pronged approach to quality-assurance**:
    1. First, with respect to data collection – as explained in section 5 – we will implement a **systematic human validation process** to double-check that social media data scraped from the internet is relevant for us. This is line with the principle 4 of 'validate, validate, validate' from Grimmer & Stewart 2013
    2. Second, all research products go through a **round of peer-review internally** (i.e. within OPM), conducted by a **senior evaluation expert**.
    3. Third, products will then be **reviewed by stakeholders**, which in this case means **UNICEF** (and UNICEF's quality assurance process) and **Government counterparts**, if possible.

9. Deliverables, Timelines and Dissemination Plan

3. The **list of stakeholders** includes **UNICEF**, **Kazakhstan**; the **Ministry of Enlightenment**, **Government of Kazakhstan** who are **the principal users of this research**. Also, stakeholders include different **organizations and NGOs** that represent or provide a voice to disabled children facing challenges in accessing inclusive education in Kazakhstan. Engagement with the last group would be mediated via UNICEF. The feedback from all the stakeholders will be incorporated into the main technical report and the capacity building workshop.

4. We will work closely with UNICEF Kazakhstan to conduct dissemination of the findings and propose at-least a round of final presentation or a joint workshop with the Working Group on Inclusive Education.

10. References and Data Sources

- Central Asian Bureau of Analytical Reporting (CABAR) (2019). One Third of Kazakhstan Schools Have Outdoor Toilets Only. https://cabar.asia/en/one-third-of-kazakhstan-schools-have-outdoor-toilets-only (Last accessed 27th March 2023).
- DataReportal, Digital 2022: Kazakhstan (2022). https://datareportal.com/reports/digital-2022-kazakhstan (Last accessed: 8th March 2023)
- Government Social Research (2016). Using social media for social research: An Introduction (Last accessed: 8th March 2023)
- Grimmer, J. and Stewart, B.M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political analysis, 21(3), 267-297
- Ministry of Education(2020). A Situational Analysis of Children in Kazakhstan. Доклад о положении детей в Республике Казахстан, 2020 год (www.gov.kz)

10. References and Data Sources

- Ministry of Education(2023). National report on the state and development of the education system KZ, 2023
- Statcounter Globalstats, Social Media Stats in Kazakhstan (2023). Social Media Stats in Kazakhstan (Last accessed: 8th February 2023)
- Sairambay, Y. (2022). Internet and social media use by young people for information about (inter) national news and politics in Russia and Kazakhstan. *Studies of Transition States and Societies*, *14*(1), 56-70
- UNESCO; UNICEF; Global Partnership for Education; (UK) Foreign, Commonwealth & Development Office, (2021) Education Sector Analysis Methodological Guidelines Volume III. Page 35
- United Nations Population Fund (UNFPA), 2020. UNFPA Казахстан | Система реагирования на насилие должна учитывать права людей с инвалидностью

Thank you

Oxford Policy Management

www.opml.co.uk

OPMglobal

11. Annex A: Our background and capabilities

1. Founded in 1979, Oxford Policy Management (OPM) is committed to helping low- and middle-income countries reduce poverty and disadvantage through public policy reforms.
2. We work in all areas of social and economic policy and governance. Our cutting-edge research capabilities and understanding of decision making and policy processes has enabled us to collaborate with many of the leading names in policy development, including UNICEF.
3. The "*Inclusive Education Evaluation by using Big Data Sources of Information*" project will be led by our Data Innovations Team (DIT), which is situated in OPM's Research and Evidence Practice in Oxford, United Kingdom (UK).

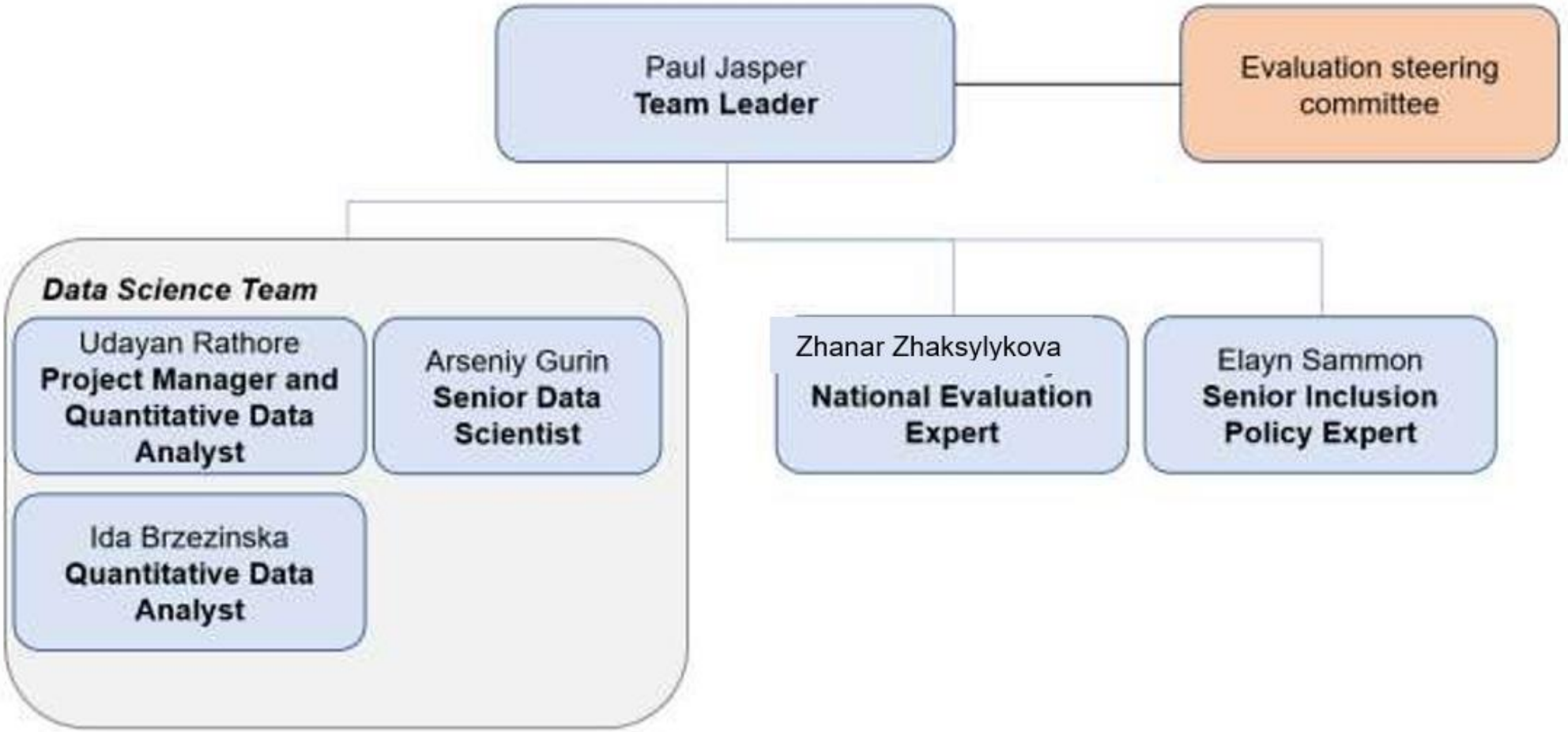
11. Annex A: Our background and capabilities

4. DIT at OPM is focused on the application of modern Data Science methods to:
   a) Support the conventional evaluations ;
   b) Provide quick and robust insights where traditional evaluations cannot be undertaken.
5. We specialize in use of innovative methods like conducting *Geospatial Analysis* and *Natural Language Processing* (NLP) to generate evidence across data cycle and support research and evaluations.
6. More information about our Data Innovation work can be found *here.*
7. A recording of our presentation on Practical Applications of NLP in Evaluations to UNICEF's Evaluation office from October 2022 can be found *here.*

11. Annex B: Key international conventions and declarations on disability and inclusive education

Box 2. Key International Reference Documents on Disability-Inclusive Education
UN Convention on the Rights of the Child, 1989 introduces the right to protection from discrimination in grounds of disability for the first time in international human rights law.
World Declaration on Education for All, 1990 highlights the steps needed to provide equal access to education to every category of disabled persons as an integral part of the education system.
UN Standard Rules on the Equalization of Opportunities for Persons with Disabilities, 1993 elaborate the steps needed to translate the principle of equal primary, secondary and tertiary educational opportunities for children, youth and adults with disabilities in integrated settings into practice.
Salamanca Statement and Framework for Action on Special Needs Education, 1994 introduces the guiding principle that ordinary schools should accommodate all children, regardless of their physical, intellectual, social, emotional, linguistic or other conditions.
UN Convention on the Rights of Persons with Disabilities (CRPD), 2006 introduces an obligation to ensure an inclusive education for persons with disabilities at all levels.
UNCRPD General Comment No. 4: Article 24: Right to Inclusive Education, 2016 elaborates the measures governments must introduce to guarantee inclusive quality education for all persons with disabilities.
SDG 4 introduces commitment to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.
Education 2030: Incheon Declaration and Framework for Action – Goal 5 aims to expand early intervention and education of children with disabilities.

11. Annex C: Team structure



11. Annex D: Team qualification and experience

Name	Position	Skill Areas											
		Graduate degree (Master's or higher)	End-user focussed data projects	UNICEF project experience	2 years project management or more	Central Asia experience	Education and/or social protection evaluation	2 years of Data Science project or more	Social Media Listening experience	Geospatial analysis experience	Kazakh language	English language	Russian language
Paul Jasper	Team Leader	•	•	•	•		•	•	•	•		•	
Arseniy Gurin	Senior Data Scientist	•	•	•	•			•	•	•		•	•
Zhanar Zhaksylykova	National Evaluation Expert	•		•	•	•	•				•	•	•
Elayn Sammon	Senior Inclusion Policy Expert	•		•	•	•	•					•	
Udayan Rathore	Project Manager and Quantitative Data Analyst	•	•	•	•		•	•				•	
Ida Brzezinska	Quantitative Data Analyst	•	•	•			•	•		•		•	

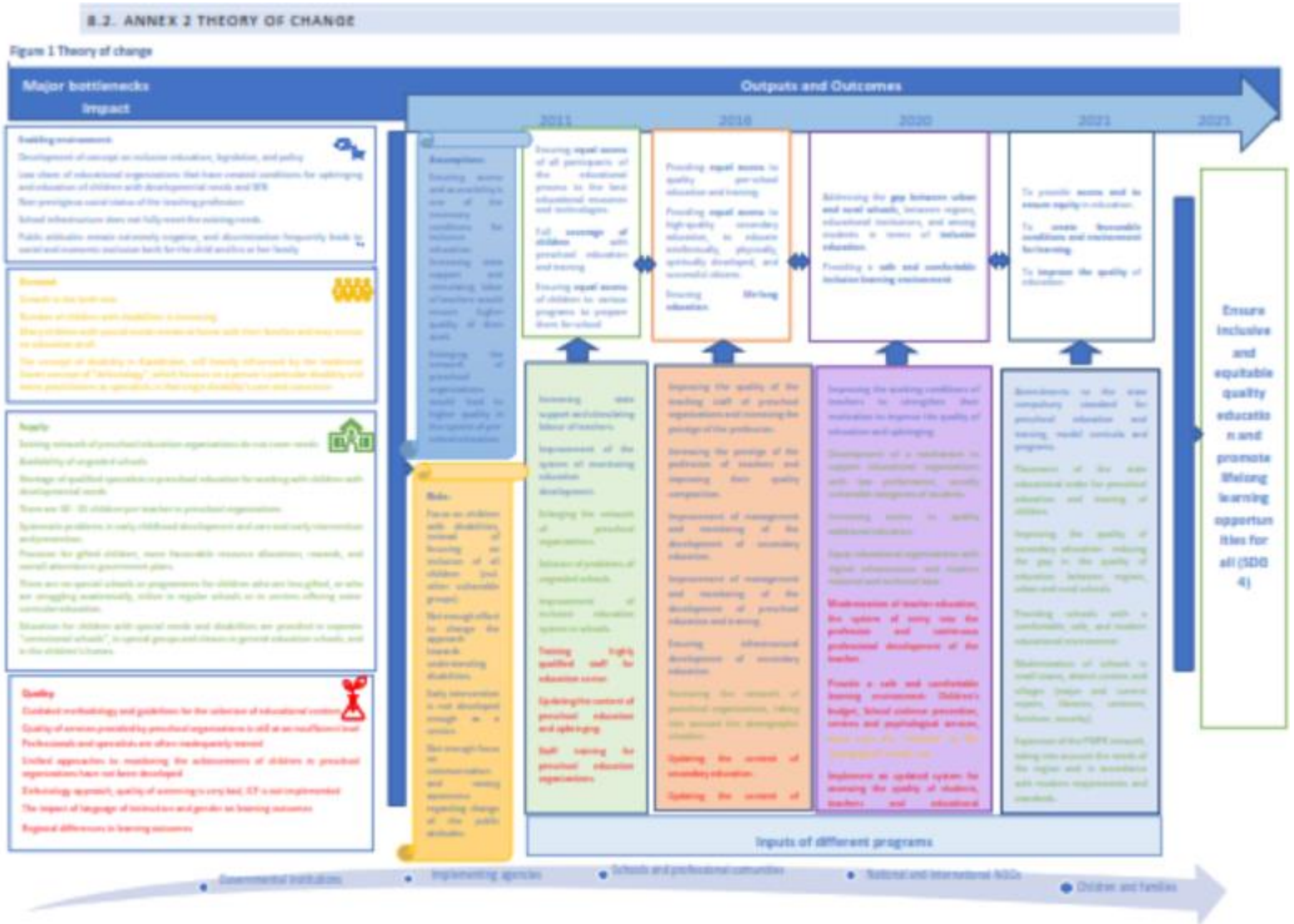
11. Annex E: Team member wise level of effort

Name	Position	Tentative level of effort on the project (invoiceable days)
Paul Jasper	Team Leader	09
Arseniy Gurin	Senior Data Scientist	17
Zhanar Zhaksylykova	National Evaluation Expert	18
Elayn Sammon	Senior Inclusion Policy Expert	08
Udayan Rathore	Project Manager and Quantitative Data Analyst	40
Ida Brzezinska	Quantitative Data Analyst	35

11. Annex F: Terms of Reference

- Please refer to the separate attachment shard with the slides.

11. Annex G: Theory of Change (Junction Bulgaria)



Reference:
Junction Bulgaria (2022). Formative Evaluation of State Programmes for Development of Education with Focus on Inclusivity. Inception Report, pp. 48.