The Labour and Welfare Bureau of the Government of the Hong Kong Special Administrative Region

Provision of Consultancy Services for Developing a Central Databank on Children
(Ref: LWB CoC/7-2/1/2)

Final Report
Executive Summary

August 2023

Table of Contents

Executive Summary

1. Introduction

1.1. Background

The Government commissioned PricewaterhouseCoopers Advisory Services Limited (Consultant) in December 2019 to assist the Working Group on Research and Public Engagement[1] (WG) of the Commission on Children (CoC) in conducting a study to explore the feasibility and implementation framework of developing a Central Databank on Children (CDC) in Hong Kong (the Study) to enable the Government and/or relevant non-government stakeholders to collect and share useful data on children.

This Executive Summary summarises key recommendations in the Final Report covering the technical architecture, governance, programme management plan, implementation roadmap of the proposed CDC model, as well as the benefits of CDC development in Hong Kong.

1.2. Literature Review of Overseas Experience

Findings from the review of five overseas CDC projects (two in the United Kingdom (UK), two in Australia and one in Canada) conducted at early stage of the Study suggested a CDC designed primarily for "trend monitoring" on the one end of the spectrum or for "case tracking" on the other end will have different implications when considering the type of data collection, which will in turn incur very different levels of data privacy concerns. For "trend monitoring", collection of aggregate and anonymised data by leveraging existing operating databases should suffice; whereas "case tracking" will require personal data which are identifiable and traceable. (See **Annex A** for summary of findings from review of overseas practices).

1.3. Local Stakeholder Engagement

Views of local stakeholders were collated through the following means.

- 20 Interviews with Government Bureaux/Departments (B/Ds), non-governmental organisations (NGOs) and the academia;
- 16 Focus Groups with 64 organisations from the social welfare sector, family groups, schools/educators as well as the social science and healthcare fields;
- 5 Engagement Sessions with over 100 attendees including children, ethnic minorities, parents of children (including parents of children with special educational needs (SEN)) and the general public; and

---

[1] Renamed to the Working Group on Research and Development since 1 January 2023.

- Over 1 000 respondents for the Survey.

  Details are set out at **Annex B**.

1.4. Broad Guiding Principles

Taking into account the literature review of overseas experience and the views collated from the local stakeholder engagement exercise, the adoption of the following guiding principles is recommended:

a) ***Purpose-specific/policy-driven:*** Overseas experience suggested that the setting of clear objectives is a key first step in the establishment of any CDC. This view is also supported by the stakeholders engaged for this Study, with many highlighting the importance of defining the objectives and priority areas for CDC development. CDC development should therefore be purpose-specific with clearly defined objectives. Due consideration should be given to its potential in contributing to the attainment of the strategic goals of CoC.

b) ***Privacy protected and secured***: A strong consensus emerging from the stakeholder engagement exercise is that privacy and data security are critical matters to be addressed in order to gain support for and public confidence in the development of a CDC. Having regard to the observed overseas experience and to address public concerns, data privacy and system security of CDC should be accorded the utmost importance when considering the type of databank and data governance mechanism for CDC development. CDC development should also take into consideration stakeholders' expectation as well as the latest development in legislation (e.g. Personal Data (Privacy) Ordinance (Cap. 486)) and measures in relation to data protection.

c) ***User-centric:*** Both findings from the review of overseas experience and views collated through the stakeholder engagement exercise highlighted the importance for CDC to adopt a user-centric design with features that best meet the needs of potential users. Where practicable, it should incorporate features that are valued by users (e.g. data discovery, data visualisation and upload/download functions etc.). In meeting the needs of key users with keen interest in children's well-being in Hong Kong, care should be taken to develop the IT framework and technical architecture of CDC so that collaboration with third parties could be realised.

d) ***Collaborative, Transparent and Consultative:*** Stakeholders generally viewed improving children's well-being as an endeavour that requires cross-sectoral and multi-disciplinary effort. To build public trust for CDC development, many highlighted the need for setting up a transparent consultative process during CDC development. This view is also in line with observed overseas practices whereby cross-sectoral collaboration and public consultation tend to be adopted when developing CDC. It is therefore suggested that CDC in Hong Kong should be developed as a multi-disciplinary and collaborative effort that involves key stakeholder groups working in sectors/fields related to children's well-being. The process of developing CDC should be transparent to the public, with the provision of a consultative channel for seeking key stakeholders' views on key areas related to its implementation.

e) ***Scalable and incremental:*** CDC development is a complex project that cuts across a wide spectrum of children-related policies/programmes under different B/Ds. As such, it will likely span across a number of years as demonstrated by observed overseas experience. In general, overseas experience suggests that it would take approximately three to six years for the deployment and development of CDC (with variation in timeline depending on factors such as the development objectives of the individual CDC or the types of data collected, etc.). It is therefore suggested that CDC should be developed in phases and in an incremental, scalable manner, including tasks and projects that can demonstrate its benefits.

f) ***Relevant and adaptable:*** As supported by findings from stakeholder engagement and overseas experience, CDC development is a sustained project rather than a time- limited exercise. There is a need to gather feedback on an ongoing basis. Development of CDC should be relevant, timely and adaptable to the evolving needs of children over time, hence the need for regular reviews to assess the development objectives and/or adjust the scope/types of data to be collected.

g) ***Net value to target beneficiaries:*** Views collected through the stakeholder engagement exercise highlighted the importance of appropriately weighing the potential benefits and costs for CDC development, particularly when determining the development objectives and appropriate types of data for collection. It is therefore suggested that the "net value to target beneficiaries" should be adopted as a key consideration in the cost-and-benefit analysis for CDC development in Hong Kong.

1.5. Key Parameters for CDC in Hong Kong

Having applied the broad guiding principles in designing the CDC in Hong Kong, the following parameters are recommended:

a) ***Development Objectives****:* The primary development objective of "trend monitoring" on the basis of non-identifiable data for the general children population should be adopted, while "prevention & early intervention" should only be adopted for specific segment(s) of the children population. In view of the concerns over data privacy collated at the stakeholder engagement exercise and the observations in **Section 1.2**, "case tracking" for the general children population should not be pursued.

b) ***Data Governance****:* While only non-identifiable data will be collected, a special procedure should be developed to safeguard personal data of the children concerned. Apart from compliance with the Personal Data (Privacy) Ordinance (Cap. 486), additional safeguards should also be put in place. This includes separate data storage and management (see elaboration in point (c) below) and institutional and operational safeguards below :

   - In the event that personal data of children will need to be de-identified and linked in data linkage projects for the purpose of realising the development objective of "prevention & early intervention", a Data Ethics and Privacy Panel could be considered for approving the purpose and usage of data collection, with amendments to be made to the Personal Information Collection Statement (PICS);

- Researchers responsible for conducting data linkage projects shall not have access to personal and identifiable data at any given point of time and measures should be put in place to ensure that data subjects will not become identifiable through reverse engineering of de-identified data; and
- There should be segregation of functions whereby different parties should be responsible for de-identifying, linking and using children data for research purpose.

An illustration of the key procedures for approving and conducting data de-identification and linkages encompassing the abovementioned safeguards is shown at **Annex C**. (See **Section 2.2** for key institutional setups responsible for the key procedures).

c) ***Data Storage and Management***: To support data linkage as mentioned above, a federated model with a two-tier database structure should be adopted. At the "macro-level", a central databank with non-identifiable data from Government bureaux and/or departments (B/Ds) will be kept for the purpose of "trend monitoring"; and at the "micro-level", the relevant B/Ds will retain data with identifiable data in their respective databanks for the purposes of "causal analysis and risk identification" (See elaboration in **Section 2.1.1**). Overall, established practice should be adopted with regard to data management and governance of CDC (e.g. the UNECE/Eurostat/OECD Generic Statistical Business Process Model (GSBPM)) to meet the specific needs of CDC operations and associated data lifecycle.

d) ***Incremental approach***: With reference to the development timeframe for CDC projects in overseas countries which took 10 to 15 years (see **Annex A**), it is recommended that the development of a CDC in Hong Kong should take about ten years in two phases. This indicative development timeframe has also taken into account the lead time required to identify appropriate set of "key data" and for B/Ds to adhere to any agreed practices/guidelines for data alignment to ensure that there would be minimal disruption to the operations of B/Ds while ensuring that new data alignment requirements will be duly considered during new development and/or revamp of IT systems. Taking into account the views expressed by stakeholders in the engagement exercise in six priority areas[2], the WG has selected "Children with Risk of Abuse & Neglect" and "Children with SEN" as the priority areas, on the consideration that: (i) the target segment can be clearly defined, and data could be relatively easy to collect and retrieve for meaningful analysis and policy formulation; (ii) relevant data of the target segment are currently captured by multiple Government departments/parties; and (iii) the potential impacts would be significant and measurable. The execution of pilot projects is also anticipated to be key in providing recommendations on data to be prioritised for alignment (See the analyses of the risks and opportunities of database design for the two priority

---

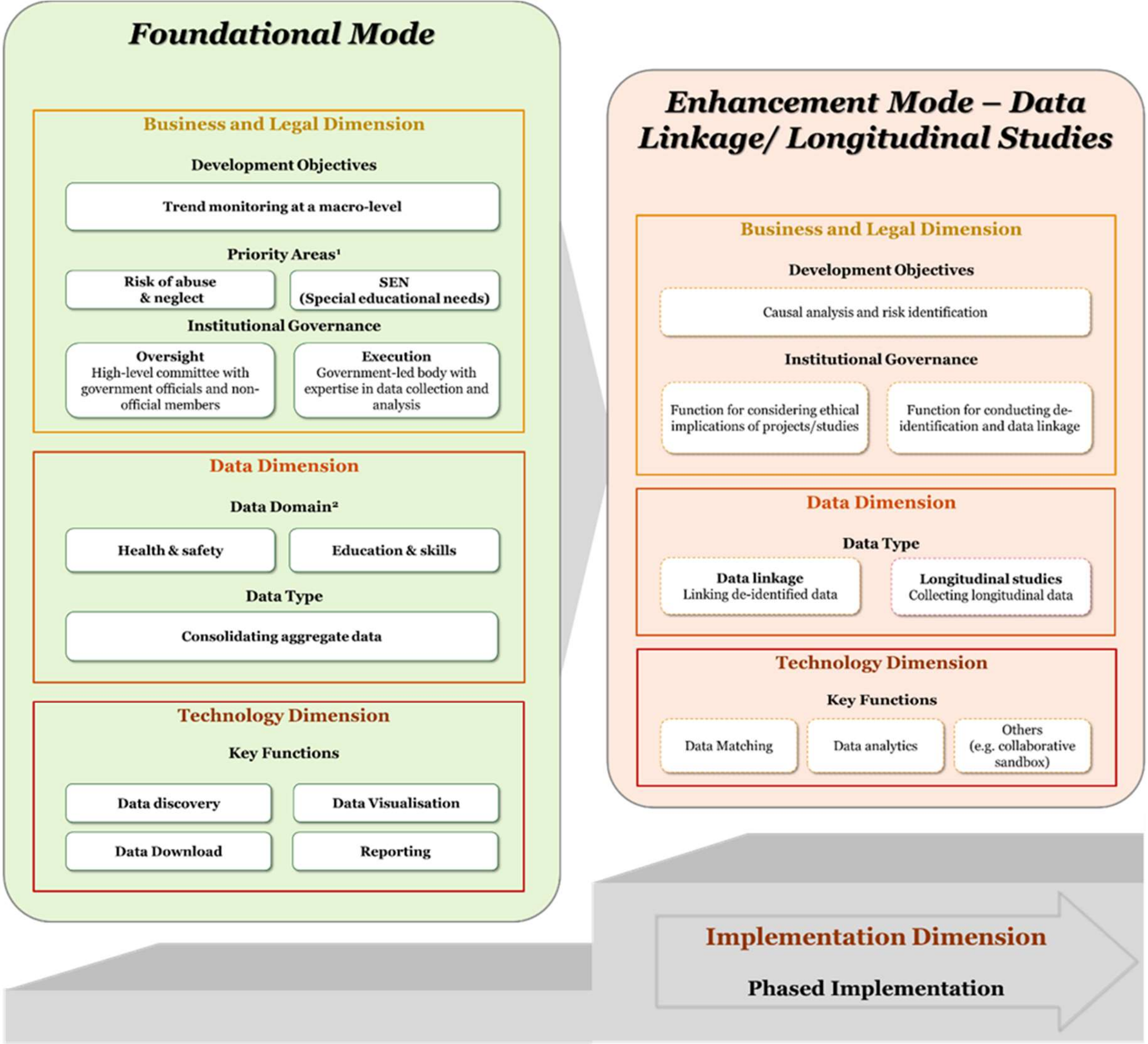
[2] The six priority areas include "Children with Risk of Abuse & Neglect", "Children with Special Educational Needs", "Chronic Health Conditions", "Living in Poverty", "Special Family Background" and "Ethnic Minorities".

areas in **Sections 2.3.4 and 2.3.5**). Key tasks to be conducted and the responsible parties involved under the incremental approach are set out in **Annex D**.

2. Proposed Central Databank in Hong Kong

Based on the abovementioned key parameters, two modes of CDC development (i.e. Foundational Mode and Enhancement Mode) have been derived to be implemented in two phases as illustrated in **Figure 1** below.

Figure 1 Overview of Key Distinctive Features for CDC Design



Note 1: For the Foundational Mode, attention shall be paid to the consolidation of relevant aggregate data across B/Ds with emphasis on macro trend monitoring of children in general and the two priority areas of “Children with Risk of Abuse and Neglect” and “Children with SEN”.

Note 2: CDC may collect existing aggregate data pertaining to any data domain at the outset of its implementation, but with a particular emphasis on “healthy and safety” and “education and skills”.

Note 3: While the Foundational Mode focuses more on data sharing among B/Ds, the Enhancement Mode allows for data sharing with external parties with an approval mechanism in place.

2.1. Technical Architecture

2.1.1. IT Framework for Foundational Mode & Enhancement Mode

An IT framework has been designed to support the federated model for CDC development with the objective of minimising disruption to various data sources arising from technical integration, and at the same time, respecting the ownership, control and confidentiality of corresponding data sources. Under this IT framework for the Foundational Mode, the proposed technical architecture would support the federated model through the setup of a two-tiered structure for the processing and dissemination of macro-level and micro-level children related data from B/Ds. Data shared by B/Ds shall be transmitted through a secure conduit over the Government Backbone Network (GNET) with encrypted Application Programming Interface (API) through a file interchange mechanism to CDC. In order to cater for the security, flexibility and volume of data to be processed, the IT framework for CDC shall also leverage the Government Cloud Infrastructure Services (GCIS) for (i) higher data, application and system security; (ii) better computing resource scalability; and (iii) more flexibility for on-demand service provision when necessary. (See **Annex E** for the proposed technical architecture under the IT framework for Foundational Mode).

For the Enhancement Mode, the two-tiered, federated model IT framework is designed to ensure that sensitive data will remain within relevant individual databases for control and ownership. Under the proposed technical architecture of CDC, sharing of micro-level data would be conducted only through an agreed data sharing mechanism initiated by CDC based on requests from CDC authorised users, and with agreement from corresponding data source owners, i.e. B/Ds, NGOs and other external organisations, such as the academia. Upon request and after internal clearance, individual data source owners would extract relevant data and share the de-identified datasets with CDC through the secure conduit over GNET with encrypted API of the file interchange mechanism to CDC. The IT framework for Enhancement mode is a full-featured design that is capable of supporting the realisation of the Enhancement Mode of CDC for both macro-level and micro-level data processing and analysis of different research projects relating to the various dimensions of children well-being. Additional data analytic capability would also be introduced through the adoption of the Government's Big Data Analytics platform. (See **Annex E** for the proposed technical architecture under the IT framework for Enhancement Mode)

In gist, the key differences between the two modes are:

- data sources of the Foundational Mode include only children related data from B/Ds for priority areas only, while that of the Enhancement Mode include children related data from B/Ds and other external organisations such as NGOs and the academia; and
- on top of the employment of statistical tools and adoption of international standards for data processing and quality assurance in the Foundational Mode, the Enhancement mode introduces a big data analytics platform; and

The differences are further elaborated at **Annex E**.

2.1.2. Potential Data Exchange with Third Parties

For the purpose of data exchange, the subscribed Shared File Interchange Service from GCIS would serve as the universal data exchange mechanism between CDC and all external data sources, including government B/Ds or other parties outside of the Government such as NGOs. The Shared File Interchange Service of GCIS would allow CDC to collect from various data sources including B/Ds or NGOs, regardless of the formats of survey data and statistical returns being transferred from the data sources to CDC, i.e. as pre-processed data files or direct data exchange files from systems, or the technical maturity and transmit protocols of the data source organisations.

2.1.3. Estimated Sizing

Based on the codebooks provided by the five B/Ds, the sizes of CDC under Foundational Mode and Enhancement Mode are estimated to be 11.9 terabytes (TB) and 67.3 TB as of 2022 respectively. The estimations are then used to tabulate the projected CDC sizing for CDC development from 2023 to 2047 assuming that the project will be undertaken for a 25-year period. The projected CDC data sizing for the Foundational Mode is 13.6 TB starting from 2023 and 23.8 TB in 2027 (Year 5 of Foundational Mode). As for the Enhancement Mode, the projected CDC data sizing is estimated to be 156 TB starting from 2028, and 2,215 TB in 2047 (Year 20 of Enhancement Mode). The projected CDC sizing for the respective modes are tabulated below:[3]

Table 1 Projected CDC Data Sizing for Foundational Mode

Projected CDC Data Sizing for Foundational Mode					
Year	2023	2024	2025	2026	2027
Size (TB)	13.6	15.7	18.0	20.7	23.8

Table 2 Projected CDC Data Sizing for Enhancement Mode

Projected CDC Data Sizing for Enhancement Mode										
Year	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037
Size (TB)	156	179	206	237	272	313	360	414	476	548

---

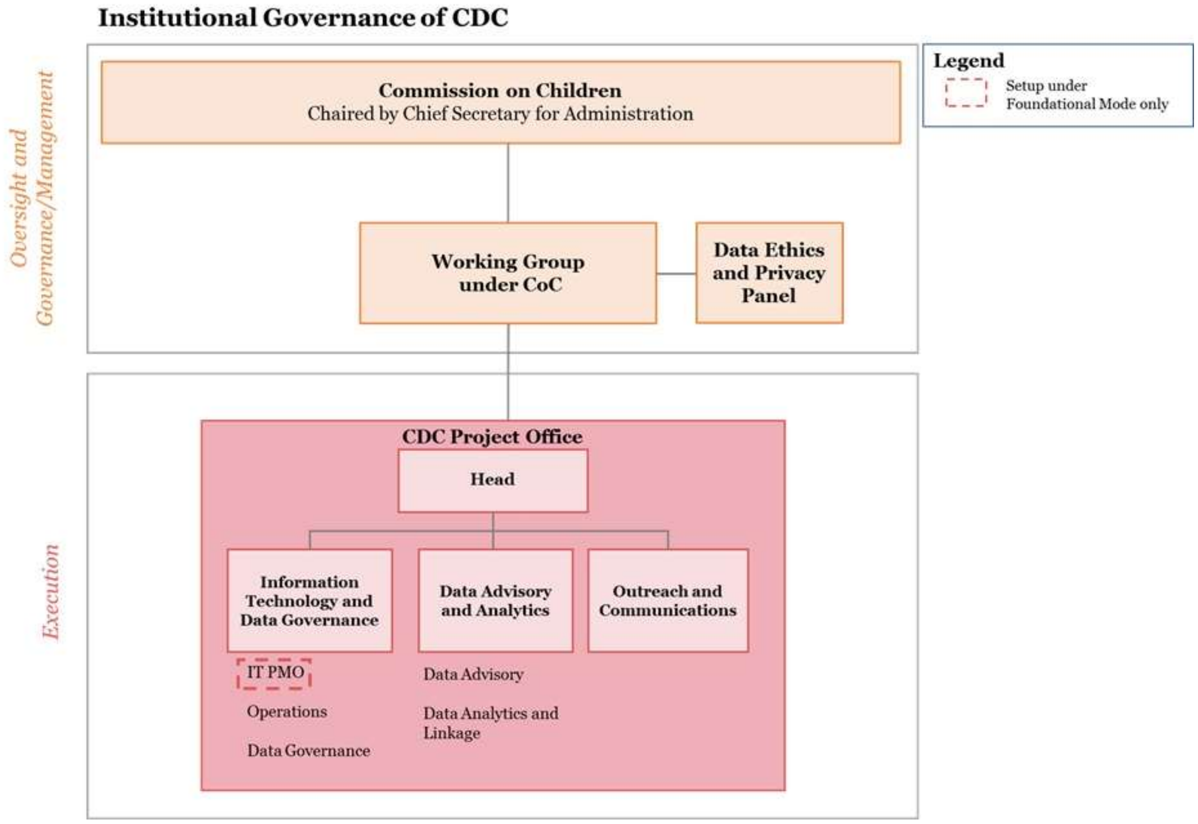
[3] For Foundational Mode, the data is expected to cover both “Health & Safety” and “Education & Skills” children well-being dimensions. For Enhancement Mode, the data is expected to cover all five children well-being dimensions, namely, “Health & Safety”, “Education & Skills”, ”Material Well-being”, “Behaviours & Risks” and “Family & Peer Relationships”.

Projected CDC Data Sizing for Enhancement Mode										
Year	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047
Size (TB)	630	724	833	958	1,101	1,267	1,457	1,675	1,926	2,215

2.2. Governance

To deliver CDC Model, a two-tiered governance structure with clear assignment of roles to entities in undertaking oversight and executive functions of CDC will be established. **Figure 2** illustrates the proposed institutional setup of CDC.

Figure 2 Proposed Institutional Setup of CDC



The potential functions for each of the proposed body are elaborated below.

Oversight Bodies

a) ***Working Group under CoC:*** It is proposed that CDC be overseen by a relevant WG under the CoC, whose functions shall include:
   - To high-level oversight for CDC development giving due consideration to the strategic priorities of children's well-being of the Government;

  - To ensure that the priority areas and scope of data inclusion in CDC development remain relevant and in line with the strategic priorities of children’s well-being; and
  - To identify key policy areas to be prioritised for data alignment and analysis, and oversee the formulation of data standards/guidelines including usage of data under the purview of research ethics with reference to local and overseas best practices.

b) ***Data Ethics and Privacy Panel:*** For the purpose of implementing the pilot projects which involve the use of de-identified data, an Interim Data Ethics and Privacy Panel comprising members with expertise in relevant areas shall be formed to provide relevant support and recommendations on data privacy and ethical issues of the pilot data linkage projects to the WG overseeing CDC. It is envisioned that a formalised Data Ethics and Privacy Panel will need to be established with a full membership for implementing the Enhancement Mode of CDC in anticipation of the increasing number of research proposals to be examined and approved on a regular basis.

Executive Body

c) ***CDC Project Office:*** A CDC Project Office within the Government will be responsible for implementing the Project Management Plan (PMP) for CDC development under the advice of the WG overseeing CDC. Relevant B/Ds, such as data contributing agencies and technical advisors, shall also be invited to provide advice on a need basis. The suggested component divisions of CDC Project Office and their key functions are as follows:

  - *Information Technology (IT) and Data Governance*: to operate and maintain CDC platform, perform functions in relation to data governance, and provide project management for the initial development of CDC;
  - *Data Advisory and Analytics*: to provide assistance on research design and data requirement to parties commissioned to conduct studies, coordinate and liaise for data alignment and conversion, and conduct data linkage; and
  - *Outreach and Communications*: to engage and conduct regular consultation with key external stakeholders, and to develop and provide training for B/Ds.

2.3. CDC Programme Management Plan

2.3.1. Compliance with Privacy Laws

The Personal Data (Privacy) Ordinance (PDPO) will be primarily the legislation that offers principal personal data protection. Under the Foundational Mode, given that only aggregate and anonymised data will be shared, there is likely to be compliance with PDPO. In addition, CDC platform shall also adopt commonly recognised standards, such as the Interoperability Framework for E-Government. Under the Enhancement Mode, given that de-identified data and longitudinal data could be shared, additional safeguards, such as the setting up of a formalised governance body for safeguarding data ethics, would need to be put in place in addition to compliance with PDPO. Both the Department of Justice and the Privacy Commissioner for

Personal Data should be consulted for a more detailed assessment on the privacy and legal implications for sharing of de-identified data, particularly for retrospective sharing. To instil public trust, it is recommended that the PICS adopted by relevant B/Ds be updated in anticipation of future data sharing opportunities with CDC.

2.3.2. Database Design for Two Priority Areas

Codebooks of relevant data variables have been obtained from five B/Ds to collect details on data in relevant databases that are relevant to the two priority areas of “Children with Risk of Abuse and Neglect” and “Children with SEN”.

It is noted that the variables of the codebooks from the relevant databases are classified into three main categories, including (i) demographic data; (ii) subject-specific data; and (iii) other children-related data. These categories could be further classified as follows:

- demographic data can be classified as: (1a) demographic data which are recorded by multiple B/Ds’ systems; and (1b) demographic data which are recorded by one B/D’s system only;
- subject-specific data can be classified as: (2a) key subject-specific data (i.e., incidence rates of different forms of child abuse & neglect) which are recorded and adopted by multiple B/Ds’ systems and the approach of which data is defined and collected is methodologically consistent; (2b) other subject-specific data which are recorded by multiple B/Ds’ systems but the approach and assessment framework might not be methodologically consistent; and (2c) other subject-specific data which are recoded by one B/D’s system only; and
- other children-related data can be classified as: (3a) other children-related data related to the policy domains of specific B/Ds’ systems but are recorded by multiple B/Ds’ systems; and (3b) other children-related data which are recoded by one B/D’s system only.

See **Annex F** for information on databases considered for formulation of data alignment.

Based on the codebooks received, some data were found to be recorded in the systems of multiple B/Ds, including (i) demographic data; (ii) four types of abuse and neglect[4]; (iii) hospitalisation[5]; (iv) 12 SEN types[6]; and (v) six other children-related data. The

---

[4] Four types of abuse and neglect (i.e. neglect, physical abuse, psychological abuse and sexual abuse) are found to be recorded by multiple B/Ds’ systems upon consolidation of codebooks of relevant databases.

[5] The variable denoting whether the child has been admitted in hospital due to the incident of abuse and neglect was found to be recorded by the systems of multiple B/Ds.

[6] 12 SEN types are found to be recorded by multiple B/Ds’ systems upon consolidation of codebooks of relevant databases, including Attention-deficit/hyperactivity disorder, Autism Spectrum Disorder , Emotional and Behavioural Difficulties, Hearing impairment, Intellectual disability , Mental illness,

consistency of data structure for children data have been examined with regard to their respective construct[7], variable types[8], values[9], and value labels[10]. It is noted that the data structures of most of these data recorded in more than one system do not align, i.e. a piece of information measuring the same aspect of a child could thus be recorded using different variable labels, variable types, values and value labels in the systems of various B/Ds. Taking demographic data as an example, while gender is recorded in all the systems examined, some systems use string to record it while other systems use numeric. Given the lower level of complexity involved in the variable of gender, it could be aligned if B/Ds which record gender using string change its corresponding variable type to numeric.

It must, however, be noted that not all data recorded in multiple systems using different data structures are suitable for alignment. For instance, although the severity level of intellectual disability is recorded in more than one system, some B/Ds record this data based on the assessment of educational psychologists whereas other B/Ds may rely on the assessments of clinical psychologists, psychiatrists, medical doctors, ophthalmologists, audiologists, or speech therapists instead. Considering the varied expertise and standards involved in the assessment process of different B/Ds, performing data alignment for this variable may require alignment of assessment tools adopted by different B/Ds. Under this circumstance, de-identified datasets from these different B/Ds should be made available to the researchers, who would subsequently select the appropriate datasets to be adopted having regard to their research objectives.

2.3.3. Data Alignment of CDC

Based on the considerations at **Sections 2.3.1 and 2.3.2**, a list of potential data items to be aligned are identified and summarised at **Annex G**.

2.3.4. Risk Analysis

In light of the above analyses and the current state of data storage in the five B/Ds concerned, the potential risks of data migration from the database of B/Ds to CDC should not be underestimated, viz:

a) B/Ds may not be able to share data due to regulatory reasons;

b) Research participants may potentially face unintended consequences, such as labelling effect, caused by data collection targeting a specific cohort of children; and

c) B/Ds may have different technical readiness in performing data sharing with CDC.

Motor impairment, Physical disability, Specific Learning Difficulties, Speech impairment, Visceral disability and chronic illness, and Visual impairment.

[7] "Construct" refers to the same aspect or dimensions that the variables are measuring.

[8] "Variable type" refers to the types of each variable, i.e. numeric, string.

[9] "Values" refers to the actual coded values in the data for each variable, i.e. 1, 2, 3.

[10] "Value labels" refers to the textual descriptions of the coded values, i.e. male, female.

Elaboration of the risk and proposed mitigation strategies are set out in **Annex H**.

2.3.5. Opportunities

The CDC platform upon data alignment can potentially contribute to facilitating research in respect of the pilot projects on "Children with Risk of Abuse & Neglect" and "Children with SEN". The indicative research questions (RQs) for the two pilot projects are set out below:

Pilot Project 1: Children with Risk of Abuse & Neglect

- RQ1 – What are the macro-trend and incidence rates of reported child abuse?
- RQ2 – What are the underlying differences for case reporting across databases from Social Welfare Department, Hospital Authority and Hong Kong Police Force?
- RQ3 – What are the risk factors of reported child abuse cases?
- RQ4 – Are there any geographical differences in distribution of reported child abuse cases?

Pilot Project 2: Children with SEN

- RQ1 – What is the number of children with SEN in Hong Kong?
- RQ2 – What are the factors contributing to differences in reporting of children with SEN conditions across various databases?
- RQ3 – What are the characteristics of children with SEN that can influence health, safety and/ or education outcomes?
- RQ4 – What is the prevalent mode and utilisation of health, educational and social service provision (i.e. across or within District Council districts / constituency areas) for children with SEN?

The anticipated relations among data sources, classification of data into broad categories and the illustration of the research topics for the two pilot projects mentioned above are set out at **Annex I**.

For the purpose of compiling a consolidated children database on the basis of the data from B/Ds' systems, three stages of data alignment detailing the general steps for alignment are proposed and tabulated below.

Table 3 Proposed Three Stages of Data Alignment

Steps	Actions	Parties Involved
Stage 1 Preparation Work		
1	The relevant WG under CoC overseeing CDC decides a priority area (or a specific topic), then the Data Advisory Unit of CDC Project Office proposes a research framework. During this step, decision should be made on whether the research project is a one-off project or a trend-monitoring one to be conducted on a regular basis.	WG, CDC Project Office

Steps	Actions	Parties Involved
2	The relevant B/Ds examine the data of the systems they managed that are related to the proposed research framework in consultation with the Data Advisory Unit of CDC Project Office; and discussions or meetings may be required with the relevant professionals in the B/Ds.	B/Ds, CDC Project Office
3	The relevant B/Ds provide a codebook for the list of variables according to the results of Step 2 to CDC.	B/Ds
4	The Data Analytics and Linkage Unit of CDC Project Office examines the codebooks, check the data structures, and classify the variables into three main categories.	CDC Project Office
5	The Data Analytics and Linkage Unit prepares a codebook of a consolidated children database and sets out the proposed data structures for the variables that are recorded by multiple B/Ds' systems after discussions with B/Ds.	CDC Project Office
Stage 2 – Data Provision and Checking		
1	B/Ds extract the relevant data.	B/Ds
2	B/Ds conduct the de-identification of the extracted data.	B/Ds
3	<u>Option 1</u> B/Ds transform the data according to the proposed data structure and submit the de-identified data to the Data Analytics and Linkage Unit of CDC Project Office. <u>Option 2</u> B/Ds submit the de-identified data to CDC Project Office, and CDC Project Office transform the data according to the proposed data structure. While there could be two options on which party to transform the data, Option 1 is preferred as B/Ds have more understanding on the data structure of their systems and some re-coding may involve sensitive data for rare child cases. Taking also into consideration the issue of data sensitivity and confidentiality, B/Ds will be in a better position to conduct the transformation process of the data alignment.	B/Ds, CDC Project Office
4	The Data Analytics and Linkage Unit checks for consistency of the demographic data provided by the B/Ds. For demographic static data, if the data are matched, new variables will be created and stored in the consolidated children database; and if the data are not matched, the Data Analytics and Linkage Unit will inform B/Ds to counter check the details. For demographic dynamic data, the latest records from the B/Ds will be retrieved and stored in the new variables of the consolidated children database.	CDC Project Office
5	B/Ds counter check the details of those unmatched demographic static data and provide the updated data to the Data Analytics and Linkage unit.	B/Ds
6	The Operations Unit of CDC Project Office updates the data in the consolidated children database. Step 4 to Step 6 should be repeated until the demographic static data are matched or in some circumstances such as	CDC Project Office, B/Ds

Steps	Actions	Parties Involved
	unavailable information and incomplete checking, consensuses have been made across B/Ds on the inconsistent demographic static data to be recorded in the consolidated children database. Discussions or meetings between B/Ds and CDC Project Office may be required for handling the data inconsistences.	
7	The Data Analytics and Linkage Unit checks for the consistency of the subject-specific and other children-related data provided by the B/Ds, and these data will be stored in consolidated children database. Where appropriate, CDC Project Office derives new variables for the subject-specific data according to various research hypotheses.	CDC Project Office
Stage 3 – Data Updating		
1	If trend monitoring is required, B/Ds will provide the data including the new entries of the children and updated dynamic data of the children who have already been in the consolidated children database to CDC Project Office on a regular basis.	B/Ds
2	The Data Analytics and Linkage Unit and B/Ds review the data structures of the variables from time to time. If required, B/Ds consider doing the adjustments on data collection or as when system will be upgraded in the future.	CDC Project Office, B/Ds
3	The Operations Unit maintains the consolidated children database.	CDC Project Office

2.3.6. Project Governance

During the implementation of CDC, the following areas should be considered as key aspects of project governance:

a) *Progress Reporting and Quality Review:* A clear line of reporting and monitoring shall be set up to ensure effective and timely communication within the various tiers of CDC's governance structure while a periodic reporting and monitoring mechanism shall also be put in place. An indicative mechanism for reporting and monitoring is proposed below:

Figure 3 Indicative Reporting and Monitoring Line for CDC

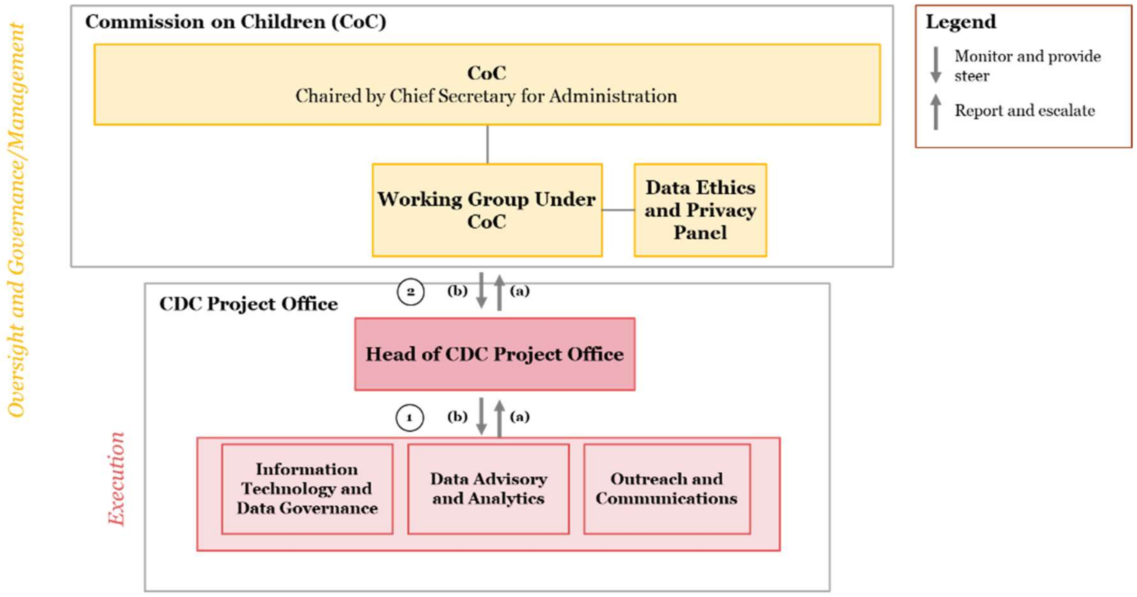
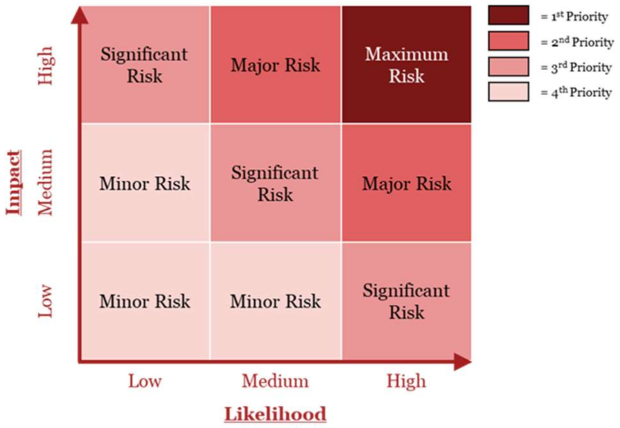


Table 4 Indicative Reporting and Monitoring Line for CDC

Item		Objective	Frequency
1	(a)	To report implementation progress and daily operational challenges encountered.	Monthly
	(b)	To manage and oversee the daily operations of the three functions within CDC Project Office, i.e.: • Information Technology and Data Governance; • Data Advisory and Analytics; and • Outreach and Communications.	
2	(a)	To report implementation progress and escalate project risks/ issues to the relevant WG under CoC overseeing CDC if required.	Bi-monthly
	(b)	To provide steer on CoC's strategic priorities for children's wellbeing and other Government child-related policies/initiatives;	
		To provide steer on cross-B/D collaboration of CDC such as: • Data alignment and data standards to be adopted by CDC; and • Data sharing to ensure compliance with PDPO and other relevant Government standards; and To endorse and approve future development plans for CDC.	Quarterly

b) *Risk Management*: Effective risk management is also a key element to the ultimate success of CDC implementation. It is important to identify, analyse and prioritise risk events that may occur during implementation. It is recommended that the risk mitigation prioritisation matrix (as illustrated at **Figure 4**) be adopted for classifying risks during CDC implementation.

Figure 4 Risk Mitigation Prioritisation Matrix



Impact \ Likelihood	Low	Medium	High
High	Significant Risk	Major Risk	Maximum Risk
Medium	Minor Risk	Significant Risk	Major Risk
Low	Minor Risk	Minor Risk	Significant Risk

Issues classified as "Maximum Risk" and "Major Risk" should be escalated to the WG under CoC overseeing CDC, whereas "Significant Risk" should be escalated to CDC Project Office Head, and "Minor Risk" to the divisional leaders of CDC Project Office. Details of selected potential risks belonging to the priority of "Major Risk" and the corresponding mitigation strategies are at **Annex H**.

c) *Critical Success Factors and Project Constraints*: A set of Critical Success Factors (CSF) (i.e. areas crucial and necessary for the project's success) for CDC development are identified at **Table 5**. In addition, with CDC being a long-term project, project constraints should also be taken into consideration to ensure the project is executed with quality, in accordance with the schedule and within budget. For instance, as a project that will span across a long period of time, CDC development may be susceptible to issue such as changes in scope that could have implications on its schedule and cost. Clearly defining the key objective of CDC at the onset and putting in place a regular progress reporting mechanism above will be important to ensure that CDC development remains to be within scope. At the same time, the mechanism should also provide flexibility for CDC to take into consideration new IT development or trends in children issues over time.

Table 5 Critical Success Factors for CDC

Areas	Critical Success Factors (CSFs)
Governance	• An established mechanism for reporting and review; and • Timely risk escalation within CDC Project Office as well as between CDC Project Office and the WG under CoC overseeing CDC.

Areas	Critical Success Factors (CSFs)
Expertise and Capabilities	• Expertise of members of the oversight bodies in relevant subject areas (e.g. children's well-being, data analytics, data standards); • Skills and expertise of the staff of CDC Project Office (e.g. data governance, research, statistical analyses); and • Capability to create a de-identification key for facilitating de-identification of data for pilot and linkage projects.
Internal Collaboration	• Effective communications on involvement required with data-contributing B/Ds; and • Close and constant liaison with and provision of support to B/Ds in relation to CDC development.
External Buy-in	• Effective communication with the public on the purposes and development progress of CDC; • Satisfactory completion of Privacy Impact Assessment (PIA) to instil public trust; and • Successful implementation of and insights from pilot projects to demonstrate the benefits of CDC.
Privacy and Ethics	• Compliance with PDPO and effective execution of safeguards for privacy protection and ethical data usage.

2.4. Implementation Roadmap

Based on the development framework of CDC, a high-level implementation roadmap for developing a CDC in Hong Kong by two phases (i.e. the Foundational Mode and Enhancement Mode) has been devised. Having regard to overseas experience and the need for data alignment among B/Ds' system in Hong Kong, it is anticipated that the Foundational Mode will be implemented over a period of five year, whereas the Enhancement Mode will be implemented over a period of ten years, on the assumption that the pre-requisites tabulated at **Annex J** will be fulfilled by the first quarter of Year 6.

2.4.1. Pilot Projects

To achieve the contribution to CDC Platform upon data alignment as illustrated in **Annex I**, pilot projects for the two priority areas of "Children with Risk of Abuse & Neglect" and "Children with SEN" should be conducted. The expected duration for each of the pilot project is around eighteen months. As the experience of the first pilot project will benefit the implementation of the second one by avoiding possible abortive work, it is proposed that the two pilot projects should be taken forward sequentially. Hence, it will take about 36 months to complete the two pilot projects. It is also suggested that a two-stage approach be adopted for the implementation of each of the pilot, namely (i) Stage 1 - Inception and Data Matching; and (ii) Stage 2 - Data Analysis and Reporting.

2.4.2. Stakeholder Engagement

To ensure CDC development is fit-for-purpose and to build public trust in CDC, the importance of having targeted and on-going stakeholder engagement was emphasised in our findings from overseas research and stakeholder consultation. The objective of adopting a comprehensive stakeholder engagement plan is to allow appropriate information to be disseminated to various stakeholder groups and strengthen transparency and visibility of CDC development across these stakeholder groups. An indicative stakeholder engagement plan is proposed at **Annex K**. It is recommended that the stakeholder engagement plan should dovetail with the first phase of developing the Foundational Mode in the first five years, during which efforts should be made to buy in stakeholders' support for the Enhancement Mode as well.

2.5. Qualitative Benefits for CDC's Implementation

With reference to stakeholder engagement findings, overseas experience and further desktop research, it is anticipated that the development of the initial CDC model has the potential to bring about the following direct benefits to key users:

a) ***Provide an integrated and comprehensive view on current state of children's well-being:*** Provision of aggregate data under the Foundational Mode of CDC could provide a strong foundation for generating a holistic view on the state of children's well-being in Hong Kong. In the case of Children's Headline Indictors (CHI) in Australia, its data (e.g. overweight and obesity, child abuse and neglect, social and emotional well-being, etc.) was used by the Department of Health in their comprehensive review of the state of children's health and well-being in Australia during the formulation of the National Action Plan for the Health of Children and Young People 2020-2030. Furthermore, the use of a single source of data could facilitate benchmarking, with a view to understanding and identifying areas of variability. Aggregate statistics available in CDC could serve as territory-wide baselines, of which reference could be made in benchmarking within Hong Kong (e.g. between districts, segments of children) or with oversea jurisdiction.

b) ***Raise awareness of emerging trends on child-related issues*****:** Reporting of aggregate data for monitoring macro-trends under the Foundational Mode has the potential to highlight existing and/or emerging issues that require further investigation. The demonstration of territory-wide children's statistics with the availability of further data breakdown could also bring awareness on challenges experienced by specific segments of children. In the case of Australia Well-being Monitoring Framework (WMF), the finding that Aboriginal young offenders were 20 times more likely to be in detention than non-Aboriginal young people in 2012 highlights a differential treatment of Aboriginal youth in the criminal justice system.

c) ***Enhance efficiency in the search for credible and reliable aggregate data*****:** Through the provision of a data catalogue for enhanced data discovery under the Foundational Mode, there is potential for CDC development to reduce the search time for credible and reliable children's data in areas related to "health and safety" and "education and skills". It was observed in the European Union (EU) that the benefits of open data include time savings for users of applications using open data[11]. In particular, open data could improve the speed at which research can be conducted and disseminated.

d) ***Enable better design of measures for tracking children well-being in the future*****:** CDC's function of macro-level trend monitoring could allow Government B/Ds a better understanding of the trends concerning issues related to children and a more comprehensive view of children well-being whereby different well-being domains that fall under the purview of different B/Ds are duly considered. Enhanced understanding of the current landscape and changing needs of children has the potential to facilitate B/Ds in developing more appropriate measures for tracking children well-being in the future. For example, selected indicators from Australia's CHI were used to keep track of Queensland's performance and progress towards improving outcomes as well as meeting the national goals relating to child protection system.

e) ***Facilitate policy review and service planning*****:** the Government and Children's workforce alike could benefit from better application of data, enabled by the use of CDC, in reviewing their policies and planning services.

  - Foundational Mode: Reporting of macro-trends with breakdown could enable better monitoring of trend that are of high strategic priorities, facilitate understanding on the plight of specific segments of children and serve as reference material/evidence base for policy action where relevant. For example, the statistics on the time spent by children on social network in the UK Children's Well-being Measures was used as an evidence base to support the UK government's review of existing regulatory landscape for safeguarding the well-being of children online and to bring forward an online safety legislation that covers the full range of online harms.
  - Enhancement Mode: Data linkage projects and longitudinal studies allow for more robust analyses, of which findings could be used by key users to inform resource allocation during service planning. For example:
    - *Data linkage projects*: Data linkage across otherwise siloed services can further support service design and delivery, by using population-level evidence of the need for supportive, intensive and targeted services. For example, by linking data from the South Australian Early Childhood Development Project (ECDP) (i.e. data in relation to perinatal, births registration, housing and child protection) and Child and Family Health Services (CaFHS), CaFHS were able to understand different levels of adversity and vulnerability (i.e. measured the number of socioeconomic, trauma, psychosocial and health-related risk factors) experienced by infants across South Australia. This research directly informed the development of CaFHS' new model of care and, through small-area-based adversity profiling, resource allocation from lower to higher areas of need.

---

[11] https://data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

  - *Longitudinal studies*: Longitudinal studies could also facilitate better policy review and service planning through tracking the same cohort of children over time. For instance, in Canada, a more granular set of data, made available through a five-year longitudinal study tracking a cohort of kindergarten children, allowed for a more robust analysis on the development differences amongst children in different communities of Alberta. This led to the identification of incidences whereby children living in the rural Strathcona County were experiencing greater difficulty in one or more of the developmental areas. These targeted findings directly informed the decision-making of some government bodies on facilities improvement, such as local parks and recreation departments in resource planning.

f) ***Enable the identification of risk factors for child-related issues*****:** In enabling robust analyses, data linkage projects and longitudinal studies under the Enhancement Mode of CDC could support the identification of risk factors with high level of validity and applicability for policy making. Risk factors identified for child-related issues could help researchers identify further research pathways for investigation; support the Government in their policy and investment decisions; and support children’s workforce in their identification of children requiring additional support as well as early detection and intervention.

- ***Data linkage projects*****:** By linking child protection data from ECDP and data from the Australian Early Development Census, it was found that children at age 5 who experienced higher levels of contact with the child protection system were more likely to experience developmental vulnerability on 1 or more domains (i.e. physical health and well-being, social competence, emotional maturity, language and cognitive skills and communication skills and general knowledge). Particularly noteworthy was the finding that children who have ever been notified (not screened in), and never had any more serious child protection contact, are nearly twice as likely to be developmentally vulnerable at age 5. This research suggests that any level of contact with child protection, even notifications determined not to be ‘child protection matters’, is an indicator of high risk for poor child development. The finding disputes the anecdotal evidence that many initial notifications are ‘not real’ and echoed the need for a whole-of-government coordinated response to address children protection-related concerns.
- ***Longitudinal studies*****:** Through tracking the development of a cohort of children in Alberta over the course of 5 years, the Early Child Development Mapping Project (ECMap) initiative found low socio-economic levels were associated with poor early development outcomes. Where positive socio-economic levels were present, the proportion of children doing well is greater. These findings suggest that low or less favourable socio-economic status increases the risk for experiencing poorer early childhood development outcomes.

g) ***Facilitate multi-stakeholder collaboration*****:** The set-up of a CDC offers collaborative opportunities between the Government, academia, and NGOs to increase the knowledge pool in developing tools, intervention and programmes to promote children’s well-being. Academia and NGOs use more quality data from the Government in supporting their research and services, whereas the Government could leverage additional insights and expertise from the sector in formulating applicable policies. For example:

- **Foundational Mode:** The findings of the Children's Well-being Measures published by the Office for National Statistics (ONS) of the UK government provide reference points for NGOs working on children matters in advocating the use of subjective indicators in measuring well-being of children. Selected measures formed reference materials for Coram Voice, a children's rights charity, and the School of Policy Studies of the University of Bristol to build upon their own research in developing a subjective measure (i.e. survey) for looked after children (i.e. children having been in the care of local authority for more than 24 hours).
- **Enhancement Mode:** In undertaking ECMap, Alberta Education of Canada collaborated with experienced academia, such as the Offord Centre for Child Studies at the MacMaster University in Hamilton (i.e. the creator of the Early Development Instrument (EDI)), to adopt the EDI as the key measurement tool for conducting research and preparing reports. Alberta Education also contracted work to Community-University Partnership for the Study of Children, Youth and Families (CUP), based at the University of Alberta, to leverage their expertise in research (e.g. data analysis) and community engagement. In addition, community coalitions in different zones formed by members from diverse background (e.g. healthcare, education, parents and other community members) were funded to facilitate the collection of primary research data for ECMap, define the geographically boundaries to ease operation of the project as well as to disseminate research output in the community.

It should be noted that the extent to which these direct benefits will be realised depend on the development stage of CDC. For example, the development objective "causal analysis and risk identification" which will materialise under the Enhancement Mode is likely to realise the benefit of supporting policy review and children's service planning than "trend-monitoring at a macro level" under the Foundational Mode, given the greater depth of insights that could be generated by performing data linkage and longitudinal studies.

Annex A - Summary of Findings from Review of Overseas Practices

Project	**United Kingdom – Children's Well-being Measures**	**United Kingdom – ContactPoint**	**Australia – Children's Headline Indicators (CHI)**	**Western Australia – Well-being Monitoring Framework (WMF)**	**Alberta, Canada – Early Child Development Mapping Project (ECMap)**
Key Role	Nation-wide measurement for better reporting on quality of life based on children's data. Its aim is to establish measures of national well-being that adequately reflect the needs of children. Its target segment is children aged 15 and below.	An online directory for connecting national and local sources of children's contact data. Its aim is to enhance coordination of the children workforce by sharing children's information for case tracking purposes. Its target segment is general children up to the time they reached their 18th birthday.	A nation-wide measurement for guiding strategy and policy. Its aim is to help guide and evaluate policy development by measuring progress on a set of indicators. Its target segment is children aged 12 and below.	A state-level measurement framework for guiding strategy and policy comprising three components. Its aim is to provide information on a range of indicators on the well-being of children and young people across the life course overtime. Its target segment is children aged below 18 years old.	A five-year longitudinal study focusing on early childhood. Its aim is to analyse the Early Development Instrument (EDI) in its main aspects by attempting to portray developmental outcomes of Alberta's kindergarten using national benchmarks. Its target segment is children aged 5 and below.
Development Timeframe	11 years, since conceptualisation phase in 2009	N.A. (Project was terminated upon initial implementation stage)	15 years, since conceptualisation phase in 2005	13 years, since 2007	11 years, since launch in 2009 (including 5 years for data collection followed by ongoing community engagement)

Project	**United Kingdom – Children's Well-being Measures**	**United Kingdom – ContactPoint**	**Australia – Children's Headline Indicators (CHI)**	**Western Australia – Well-being Monitoring Framework (WMF)**	**Alberta, Canada – Early Child Development Mapping Project (ECMap)**
Governance	Oversight by the board of a public body focussing on statistical reporting, with operations undertaken by an executive arm.	Oversight by a government agency with project management undertaken by an operating unit within the agency.	Oversight by a high-level strategic committee, with operations undertaken by a public body with a focus on health and welfare statistics.	Oversight by a joint standing committee, with operations undertaken by a public body with a focus on children's well- being.	Oversight by a government agency focussing on education policy, with execution of the study contracted out to researchers.
Legislation	Use of existing legislations to facilitate data sharing for research purposes.	Enactment of dedicated legislation for mandating sharing of data to ContactPoint.	Use of existing legislations to operationalise a data ethics committee for facilitating data sharing for research purposes.	Adoption of legislation specifying the Commissioner's legal obligation to monitor well-being.	Children well-being legislation provided a facilitative environment for research.

Project	**United Kingdom – Children's Well-being Measures**	**United Kingdom – ContactPoint**	**Australia – Children's Headline Indicators (CHI)**	**Western Australia – Well-being Monitoring Framework (WMF)**	**Alberta, Canada – Early Child Development Mapping Project (ECMap)**
Type of data	Public access is available to 31 headline measures across seven domains, with measures compiled mainly from secondary data from existing studies and administrative sources.	Highly restricted use of basic identifying information with data contributed by national government, local authorities and practitioners.	Public access is available to 19 headline measures across three domains with data collection from studies/administrative sources from the public sector.	Online viewing of 123 headline measures across three domains with data collection from multiple sources is permissible.	Public access to aggregated research output with micro-level data gathered through primary data collection is made available.
Implementation consideration	Phased implementation with different modes of stakeholder engagement and use of guiding principles for development of measures.	Phased implementation with use of pilots, consultation with public and cost benefit assessment.	Phased implementation with consultation of state/territory government bodies, data-related committees and experts.	Phased implementation with collaboration with research institute for measures enhancement.	Execution of a statistical project with sustained momentum through proactive community engagement.

Project	**United Kingdom – Children's Well-being Measures**	**United Kingdom – ContactPoint**	**Australia – Children's Headline Indicators (CHI)**	**Western Australia – Well-being Monitoring Framework (WMF)**	**Alberta, Canada – Early Child Development Mapping Project (ECMap)**
Achievement or outcomes	Focus on raising awareness of specific children issues, contributing to advocacy work and substantiating an integrated narrative on children.	Closure due to concerns over privacy, inappropriate data access and cost overruns.	Use to inform priority areas of strategy and monitoring performance of children services.	Use to inform differential situation between general and vulnerable groups of children during the review of justice system for youths.	Collection of micro-level data enables improvement in service provision through identification of individual specific conditioning and risk factors.

Annex B - Details of Stakeholder Engagement Exercise

Interviews

Organisations	Completed
Government Bureaux/Departments (e.g. Social Welfare Department, Education Bureau, Department of Health, Hospital Authority, etc.)	12
NGOs (e.g. Hong Kong Jockey Club, Hong Kong Society for the Protection of Children, etc.)	4
Academia/Individuals (e.g. School of Public Health/HKU, Faculty of Law/CUHK, etc.)	4
Total	**20**

Focus Groups (16 completed with 64 organisations attended)

Categories	Attended
Social Welfare, Family Groups, Schools/Educators (e.g. Hong Kong Christian Services, Heep Hong Parents' Association, etc.)	49
Social Science	8
Healthcare	7
Total	**64**

Engagement Sessions (5 completed with 112 people attended)

Participants	Attended
ES#1 Children (aged 13-17)	21
ES#2 Ethnic Minorities (aged13-17)	18
ES#3 Parents of Children (aged at or below 12)	32*
ES#4 Parents of Children with SEN aged below 18	21
ES#5 General Public (aged above 18)	20
Total:	**112**

*A total of 11 children aged between 6 to 12 participated ES#3 together with their parents.

Survey

There are 1 007 respondents for the survey.

Summary of Views Gathered from Stakeholder Engagement

This part of Annex B compiled key findings on the views and expectations of key stakeholders in relation to the feasibility of developing a CDC. Views are gathered from the engagement activities conducted in Phase 3 of the Study including interviews, focus group discussions (FG), engagement sessions (ES) and a survey targeting at members of the general public.

Key highlights of stakeholders' views on the areas of Development Objectives, Scope of Data, Public Trust Building and Data Ethics and Implementation of CDC are tabulated below –

Key Highlights of Stakeholders' Views
Development Objectives
• "Monitoring children's well-being at a macro-level" should be a development objective of CDC. • There is a need to identify more specific topics or provide breakdown of aggregate data for more meaningful macro-level analysis. • Many stakeholders agreed that "supporting causal analysis and risk identification" should be considered as a development objective. • Divergent views on the approach (i.e. Option A – Collecting data for longitudinal studies and Option B – Linking existing administrative data of children) in meeting "causal analysis and risk identification". • Whilst stakeholders could see the value in "case tracking to enhance coordination of the children's workforce", they generally agreed that this objective would trigger serious privacy concerns. • Stakeholders see the potential to track specific segments of children under limited use and with consent. • Stakeholders generally expect CDC development to bring about benefits for different key parties, i.e. the Government, academia and children's workforce, and general members of the public.
Priority Areas
• Many stakeholders were of the view that children with risk of abuse or neglect and children with SEN should be considered as priority areas. • Key stakeholders' views in relation to the two priority areas are broadly summarised below: <u>Children with Risk of Abuse or Neglect</u> o This segment is commonly highlighted by stakeholders engaged through interviews and ES.

  - Over 70% of respondents (71.3%) selected this segment as a priority area in the Survey.
  - Stakeholders engaged through different modes expressed concerns in view of the increasing occurrence of child abuse and fatal cases over the past years. A few stakeholders also opined that this trend has exacerbated amidst the ongoing COVID-19 situation.
  - Many stakeholders were of the view that there would be a need to generate deeper insights in order to identify susceptible cases and provide target support.

<u>Children with SEN</u>

  - This segment is commonly highlighted by stakeholders engaged through interviews, FG and ES.
  - More than half of respondents (56.3%) selected this segment as a priority area in the Survey.
  - Stakeholders engaged through different modes were of the view that there would be a need to focus on this segment due to the lack of data available in understanding the current situation. Current provision of services to SEN children is also perceived to be inadequate.
  - There were views that the ongoing COVID-19 situation could have a negative impact on the well-being of SEN children.
  - A few stakeholders engaged through different modes suggested prioritising certain types of SEN conditions such as Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD). However, a stakeholder from FG also cautioned that prioritising specific SEN types may lead to public perception of inequitable treatment across SEN types.

Scope of Data

<u>On data domain</u>

- Stakeholders generally agreed that there would be a need to focus on more than one data domain in understanding children’s well-being, with “health and safety” and “education and skills” most mentioned.

<u>On data type and data sharing arrangement</u>

- Stakeholders generally agreed that aggregate data could be shared through CDC, with sharing of essential identifiable data triggering the most privacy concerns. There were more divergent views in the sharing of other forms of data.

<u>On data governance</u>

- Stakeholders generally agreed on the need for multi-level safeguards for protecting the privacy and security of data.

On CDC architecture and design

- A decentralised architecture for CDC development is perceived to be least disruptive to database owners.
- Many stakeholders envisioned CDC to be a platform where children-related data, information and analysis would be available.
- Several stakeholders made suggestion on the features and technology to be adopted by CDC.
- Stakeholders generally agreed that existing official data from relevant B/Ds could serve as key sources of data for CDC, although other sources of data (i.e. collected through both primary and secondary means) should also be considered.
- Stakeholders generally agreed on the need to set access rights for different stakeholder groups.

Public Trust Building and Data Ethics

On public trust building

- Many stakeholders highlighted data privacy and security as key public concerns in CDC development.
- Stakeholders generally agreed that public engagement is a pre-requisite for instilling public trust in CDC implementation, and recognised the below as key stakeholders –
    - Children workforce (e.g. social workers, educators and healthcare practitioners)
    - NGO service providers
    - Academics and experts
    - Parents/guardians
    - Children
    - Public
- Several stakeholders suggested on ways to improve public perception of CDC development including engaging public figure(s) with positive image and engaging public relations specialists for public communications.

On data ethics

- When compared with concerns over data privacy and security, very few stakeholders emphasised data ethics as a prioritised concern.
- Key suggestions on considerations for data ethics are broadly summarised below:

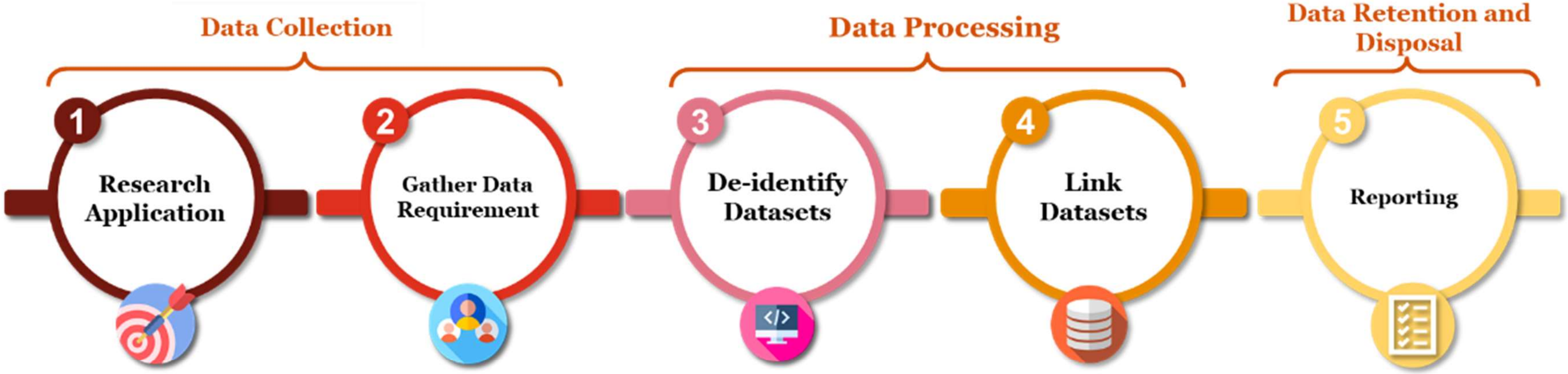
  - Formulating a clear data ethics framework governing data usage (e.g. setting principles and guidelines for use);
  - Communicating the data ethics framework to the public to gain public trust;
  - Establishing a data ethics committee (comprising independent parties with no conflict of interest and with relevant subject matter knowledge);
  - Establishing a set of protocols for assessing study designs, method of collecting data and usage of data;
  - Consulting the Privacy Commissioner for Personal Data for specific advice pertaining to ethical and privacy issues; and
  - Establishing safeguards such as the need for pre-requisite consent, setting access controls and adopting an audit mechanism (refer to "Data Governance" above).

Implementation Approach

- Stakeholders generally agreed that a phased approach for CDC implementation should be adopted although considerations would need to be given on a publicly acceptable timeline.
- Key suggestions regarding implementation priority are broadly summarised below:
  - Need to define objectives, priority areas and assess privacy implications before undertaking other tasks of designing and implementing CDC;
  - Potential to conduct trend analysis derived from aggregate and anonymised data as a foundation before embarking on research of specific topics;
  - Need to implement a pilot phase with development of use cases to demonstrate benefits of CDC with prioritisation of specific segments of children such as children with risk of abuse or neglect and children with SEN;
  - Consider aggregating more readily available data first (i.e. suggestion that health and education data would be more ready) before release of other data that is more sensitive (e.g. further breakdowns or identifiable data) or require consent for further collection;
  - Consider classifying existing data into categories to better organise data for discovery;
  - Consider consolidation or linkage of Government-owned data first before extending to other sectors;
  - Consider involvement of key B/Ds that hold the most relevant children-related data within the Government first; and
  - Consider setting up an online platform for public communications during the pilot phase.
- Several stakeholders provided suggestions on pilots to be considered for demonstrating the benefits of CDC, such as:

- Release of thematic reports on children in general or specific segments (e.g. profile of child mortality in Hong Kong, current state of children with SEN, impact of COVID-19 and social distancing measures on mental and physical wellness of children in general or on specific segments of children); and
- Release territory-wide child development data (e.g. aggregate data on child development in early years) for researcher to benchmark/reference with overseas jurisdictions.

Annex C - Approval Process for Data De-identification and Data Processing



Data Collection

1. **Research Application**
2. **Gather Data Requirement**

Data Processing

3. **De-identify Datasets**
4. **Link Datasets**

Data Retention and Disposal

5. **Reporting**

	1 Research Application	2 Gather Data Requirement	3 De-identify Datasets	4 Link Datasets	5 Reporting
Key Actions	• *Researchers* to submit research applications • *Data Ethics and Privacy Panel (Panel)* to examine the projects and approve data collection • *Researchers* to sign undertaking form	• *CDC Project Office (Data Advisory Unit)* to gather data requirements in coordination with the researcher and *data contributing B/Ds (B/Ds)*. • *B/Ds* to provide final approval for dataset	• *CDC Project Office (Data Governance Unit)* to generate the de-identification key • *B/Ds* to extract relevant data and apply the key for de-identification	• *B/Ds* to share the de-identified data with CDC Project Office *(Data Analytics and Linkage Unit)* • *Data Analytics and Linkage Unit* to merge the de-identified data and conduct consistency checks • B/Ds to follow-up on clarifications if needed	• *CDC Project Office (Data Analytics and Linkage Unit)* to pass the merged de-identified data to researchers for analysis • *Researchers* to adhere to the disposal schedule and provide notification of disposal
For Considera-tion	• *B/Ds* to update Personal Information Collection Statement				

Separation of duties between Data Governance Unit (key-holder) and Data Analytics and Linkage Unit.

Annex D - Incremental Approach for Data Alignment

Short Term - Key Objectives**:**

(1) To formalise a mechanism for identifying potential "key data" for alignment;

(2) To commence the process of setting data standards/guidelines for most prioritised "key data"; and

(3) To advise on data conversion function.

Key Tasks		Key Driver
1.	Assign a relevant WG under CoC to provide oversight to CDC, including advising on "key policy areas" to be prioritised for data alignment and analysis, and overseeing the setting of data standards and guidelines for alignment (See functions of the concerned WG at Section 2.2)	CoC
2.	Commission the pilot project for the first priority area of "Children with Risk of Abuse and Neglect", in which the engaged parties will be required to provide recommendations on data to be prioritised for alignment	CoC
3.	Identify "key policy areas" to be prioritised for data alignment and analysis taking into consideration the strategic priorities of CoC, and recommendations arising from the first pilot project	Relevant CoC WG
4.	Oversee the formulation of data standards/guideline including usage of data under the purview of research ethics with reference to local and overseas best practices	Relevant CoC WG
5.	Commission the pilot project for the second priority area of "Children with Special Educational Needs"	CoC
6.	Perform task 3 and 4 by applying the experience of the first pilot project	Relevant CoC WG
7.	Formulate data standards/guidelines for the first set of "key data" in consultation with B/Ds	CDC Project Office
8.	Provide data alignment/conversion support for pilot projects in the priority areas, and document any agreed data alignment practices among B/Ds over time	CDC Project Office

Medium to Long Term – Key Objectives:

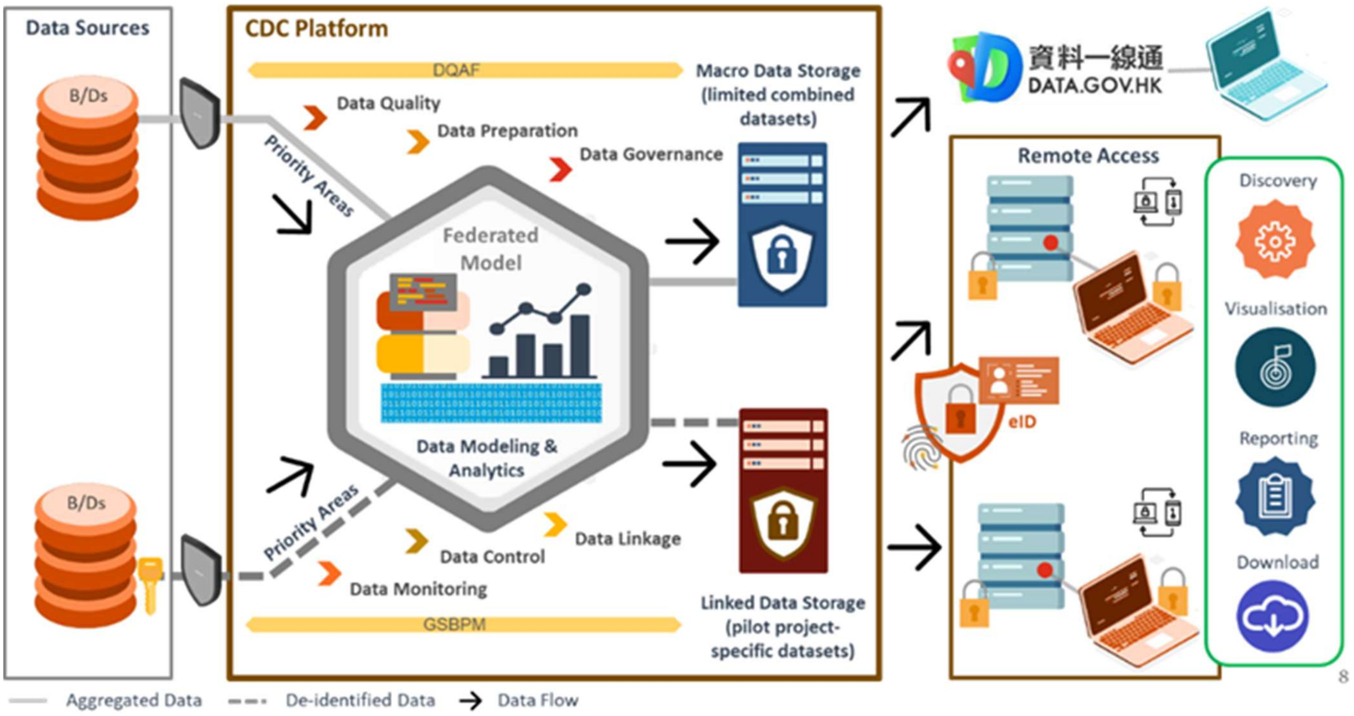
(1) To monitor the progress of data alignment and;

(2) To provide continued support for data conversion and alignment

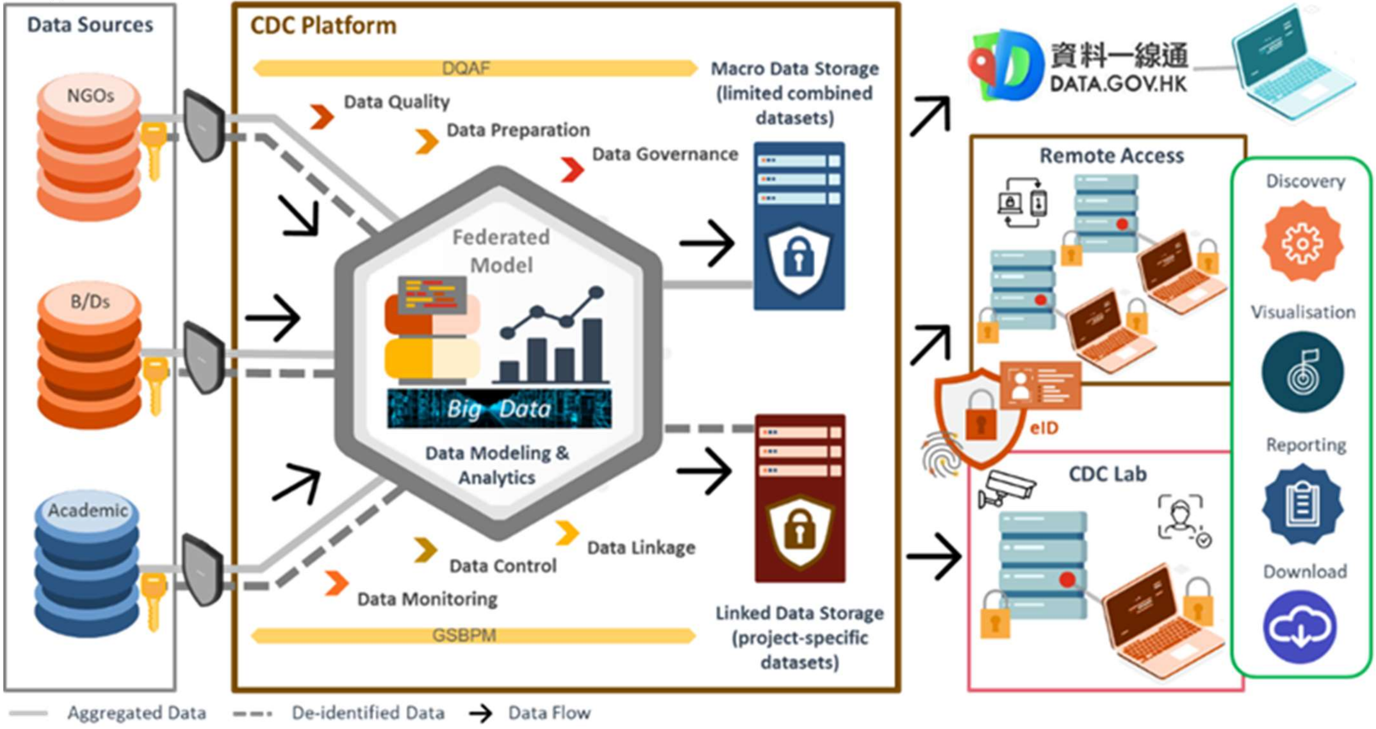
Key Tasks		Key Driver
1.	Monitor the progress of data alignment and overseeing the formulation of any new standards/guidelines such as project-specific practices/procedures for topics of concern that are investigated over time	CDC Project Office
2.	Continue to provide data alignment/conversion support for data linkage projects and provide reporting to the relevant WG overseeing CDC on potential practices that could be further aligned for specific topics of concern	CDC Project Office
3.	Identify new policy areas for data alignment and analysis, and advise B/Ds on considering new data alignment requirements with reference to data alignment exercises conducted during new development and/or revamp of BDs' IT systems where appropriate	Relevant CoC WG

Annex E - Proposed Technical Architecture under the IT Framework for Foundational Mode and Enhancement Mode

Proposed Technical Architecture under the IT Framework for Foundational Mode



Proposed Technical Architecture under the IT Framework for Enhancement Mode



Key Differences between the Foundational Mode and Enhancement Mode

Areas	Foundational Mode	Enhancement Mode
Data Sources	Children related data from B/Ds for priority areas only	Children related data from B/Ds, NGOs and other external organisations, such as the academia for various dimensions of children welfare
Data Processing	Statistical tools and adoption of international standards for data processing & quality assurance	Big data analytics platform, statistical tools and adoption of international standards for data processing & quality assurance
Data Dissemination	Enquiries by B/Ds via remote access through GNET with multi-factor authentication	Enquiries of sensitive de-identified data by B/Ds via remote access through GNET with multi-factor authentication. The IT framework of the Enhancement Mode also possess the capability for such enquiries to be disseminated in a secured environment for persons with clearance (e.g. B/Ds staff and external researchers).
Storage Capacity	Long-term storage of processed data of priority areas from B/Ds	Long-term storage of processed data from B/Ds, NGOs and other external organisations
Connectivity	Secure conduit via GNET only for data sourcing, enquiry and dissemination	Secure conduit via both GNET and Internet for data sourcing, and GNET only for data enquiry and dissemination

Annex F - Information on Databases Considered for Formulation of Data Alignment

Department Name	Database Name	Scope
Department of Health (DH)	Child Assessment Service Information System (DH-CASIS)	DH-CASIS which is operated by the Specialised Services Branch – Child Assessment Service (CAS) records the information of children who are under 12 years of age with developmental- behavioural problems or disorders.
	Child Health Service System (DH-CHSS)	DH-CHSS which is managed by the Family and Student Health Branch – Family Health Service (FHS) records the information of babies and children from birth to five years.
	System for Managing the Assessment of Student Health (DH-SMASH)	DH-SMASH which is operated by the Family and Student Health Branch – Student Health Service (SHS) records the information of primary and secondary school students.
Education Bureau (EDB)	Kindergarten Education Scheme System (EDB-KGESS)	EDB-KGESS which is managed by the School Development and Administration Branch – Kindergarten Education Division (KGED) records the information of children attending the kindergartens joining the Kindergarten Education Scheme.
	Special Education Management Information System (EDB-SEMIS)	EDB-SEMIS which is operated by the Professional Development & Special Education Branch – Special Education Division (SED) records the information of students with special educational needs (SEN) and the services include Educational Audiology Service, School-based Speech Therapy Services, School-based Educational Psychology Service, teacher professional development on catering for students with SEN as well as referral and placement services for aided special schools.
	Student Information Management System (EDB-STIMS)	EDB-STIMS which is managed by the Planning, Infrastructure and School Places Allocation Branch – Education Commission and Planning Division (ECP) records the information of students.
Social Welfare Department (SWD)	Child Protection Registry (SWD-CPR)	SWD-CPR maintained by the SWD collects and compiles statistical information on the children who have been maltreated/might have been maltreated or are currently at risk of maltreatment and the perpetrators/ alleged perpetrators/ potential perpetrators for the purpose of ascertaining the magnitude of the problem, including identification of the general profile and characteristics of child maltreatment.
	Central Referral System for Residential Child Care Services (SWD-CRSRC)	The SWD-CRSRC is maintained by SWD to record the demographic data of children who have been waitlisted/are waitlisting for residential child care services, including personal particulars, family and housing type, schooling and health condition, etc.

Hospital Authority (HA)	Cancer Statistics Query System – Children and Adolescents (HA-CanSQS).	Hong Kong Cancer Registry (HKCaR) is a population-based registry committed to collecting and conducting analyses on data from all cancer cases in Hong Kong and is responsible for managing the database of HA-CanSQS – Children and Adolescents.
	Clinical Management System (CMS) / clinical database	It records the clinical data of the patient from all public hospitals including inpatient, outpatient and emergency department admissions information.
Hong Kong Police Force (HKPF)	Case Management and Investigation System (HKPF-CMIS)	HKPF-CMIS is managed by the Information Systems Wing, HKPF. It records case information including but not limited to those related to family violence, domestic violence, sexual violence, child abuse, elder abuse and juvenile crime.

Annex G - List of Potential Data Items to Be Aligned

Data Item to Be Aligned	B/Ds Involved	Data Alignment
(1) Demographic data		
(1a) Demographic data recorded in multiple B/Ds' systems		
Age	DH, SWD	Data is aligned and recorded in numeric type (0-99)
Date of birth	DH, EDB, HA, HKPF, SWD	Data is transformed to string type in date format - DDMMYYYY
Ethnicity	EDB, SWD	Data is transformed to numeric type (1-14)
Gender	DH, EDB, HA, HKPF, SWD	Data is transformed to numeric type (1-3)
District of Residence	EDB, HA, SWD	Data is transformed to numeric type (1-20)
(2) Subject-specific data – Child Abuse		
(2a) Key subject-specific data – Child Abuse		
Neglect	DH, HA(*), SWD	Data is transformed to numeric type (0,1)
Physical abuse	HA(*), HKPF, SWD	
Psychological abuse	HA(*), SWD	
Sexual abuse	DH, HA(*), HKPF, SWD	
(2b) Other subject-specific data recorded in multiple B/Ds' system - Child Abuse		
Hospitalisation	HKPF, SWD	Data is transformed to numeric type (0,1)
(2) Subject-specific data – SEN		
(2a) Key subject-specific data – SEN		
Attention-deficit/hyperactivity disorder (ADHD)	DH, EDB, HA(*), SWD	Data is transformed to numeric type (0,1)
Autism Spectrum Disorder	DH, EDB, HA(*), SWD	
Emotional and Behavioural Difficulties	DH, HA(*), SWD	
Hearing impairment	DH, EDB, HA(*), SWD	
Intellectual disability	DH, EDB, HA(*), SWD	
Mental illness	DH, EDB, HA(*), SWD	
Motor impairment	DH, HA(*)	
Physical disability	EDB, HA(*), SWD	

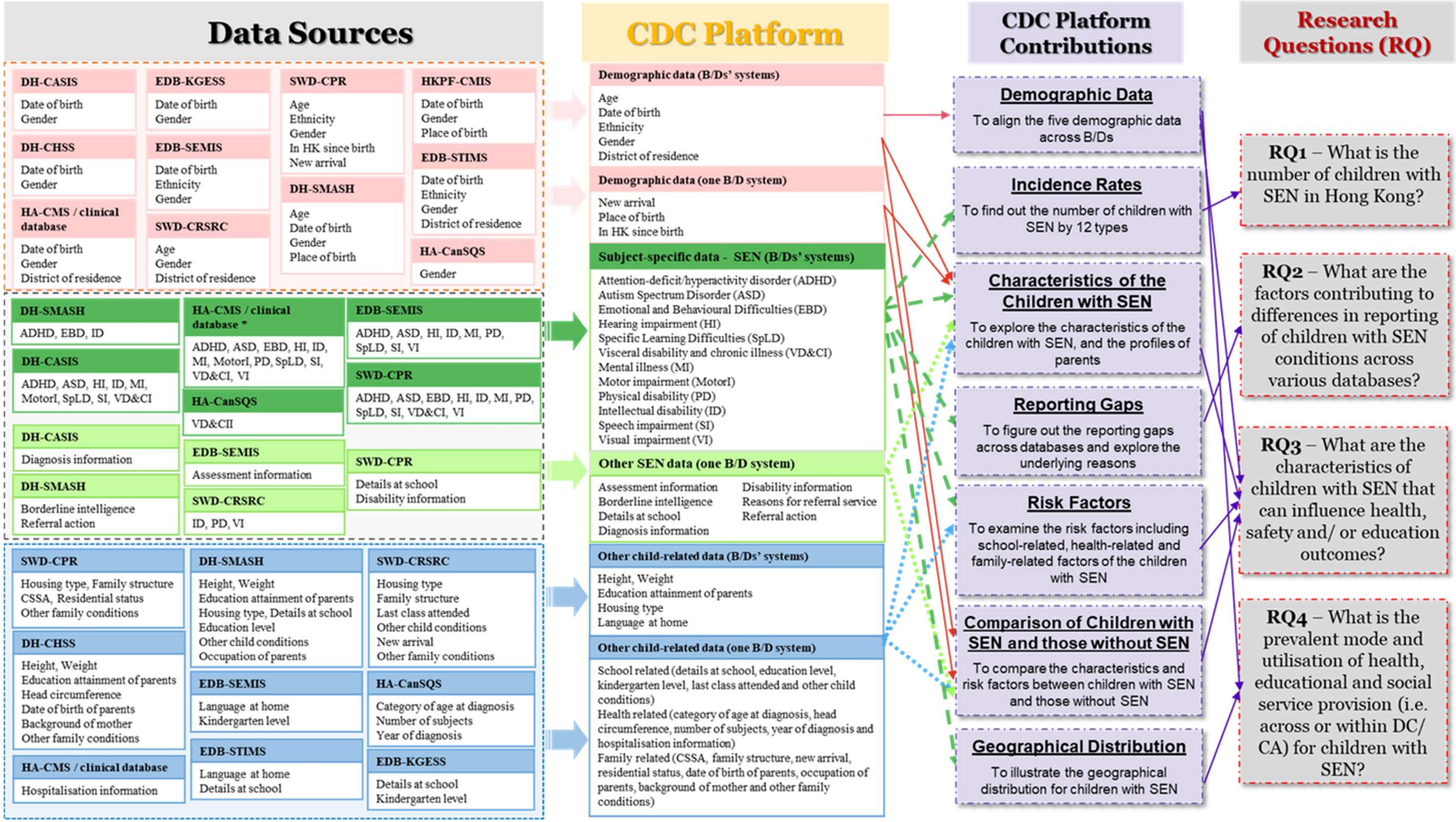
Data Item to Be Aligned	B/Ds Involved	Data Alignment
Specific Learning Difficulties (SpLD)	DH, EDB,HA(*), SWD	
Speech impairment	DH, EDB, HA(*), SWD	
Visceral disability and chronic illness	DH, HA(*), SWD	
Visual impairment	DH, EDB, HA(*), SWD	
(3) Other child-related data		
(3a) Other child-related data recorded in multiple B/Ds' systems		
Height	DH	Data is recorded in numeric type (in cm). No data alignment is required.
Weight	DH	Data is recorded in numeric type (in kg). No data alignment is required.
Education attainment of Father	DH	Data is transformed to numeric type (1-8). For DH systems - to change the variable type to numeric.
Education attainment of Mother	DH	Data is transformed to numeric type (1-8). For DH systems - change the variable type to numeric.
Housing Type	DH, SWD	Data is transformed to numeric type (1-4). For DH system and SWD systems to combine the detailed codes (say Home Ownership Scheme and Tenants Purchase Scheme into subsidised housing) into the 4 proposed codes.
Language at Home	EDB	Data is recorded in string type. No data alignment is required.

Note: For items marked with * (i.e. HA CMS / clinical database) – the diagnoses are based on ICD-9-CM, with mapping to 2010 v ICD-10. Review on the diagnosis list by subject experts, which could take up to a few months, is recommended for a more comprehensive diagnosis list for child abuse & neglect, and SEN. The number of variables in the data category are therefore to be determined after the expert review.

Annex H - Selected Potential Risks and Corresponding Mitigation Strategies of CDC Development

<table>
<tr><th>#</th><th>Potential risk</th><th colspan="3">Anticipated likelihood and impact</th><th>Mitigation strategy(ies)</th></tr>
<tr><td rowspan="3">1.</td><td rowspan="3">B/Ds are not able to share data due to regulatory reasons.</td><td colspan="3">Existing regulations (e.g. PDPO and departmental guidelines etc.) may restrict data sharing from certain B/Ds (e.g. HKPF), especially on data pertaining to individual child in child abuse cases; therefore, the likelihood of such a risk is considered to be high. Restricted data sharing from certain B/Ds may hinder certain data linkage projects or longitudinal studies that aim to discover risk factors for child abuse cases; thus, the impact of this risk is considered medium.</td><td rowspan="3">• Need for an overall policy and/or legislative support for enabling sharing of data pertaining to children with risk of abuse or neglect
• Potential for exemption from PDPO if the identity of child abuse victims are anonymised before sharing to CDC, so as to protect privacy of children with risk of abuse or neglect</td></tr>
<tr><td>Likelihood</td><td>Impact</td><td>Priority</td></tr>
<tr><td>High</td><td>Medium</td><td>Major Risk</td></tr>
<tr><td rowspan="3">2.</td><td rowspan="3">Research participants may potentially face unintended consequences, such as labelling effect, caused by data collection targeting a specific cohort of children.</td><td colspan="3">Thematic research, especially longitudinal studies on a specific cohort of children in need, may lead to unintended labelling effect on a certain group of children if the intended research purpose is not clearly conveyed to the data subjects and the public. This could lower public confidence on sharing data and trust in CDC, potentially leading the execution difficulty and suspension of specific CDC projects and impairing the image of CDC. In this regard, the likelihood of such a risk is considered medium, while its impact is considered high.</td><td rowspan="3">• Setup of multiple transparent and effective communication channels for CDC to explicitly convey its research purpose and demonstrate the benefits of such studies (e.g. development of associated support programs/policies) to the public to build rapport
• Ensure informed consent from data subjects and their parents are obtained and the possible impacts are clearly communicated</td></tr>
<tr><td>Likelihood</td><td>Impact</td><td>Priority</td></tr>
<tr><td>Medium</td><td>High</td><td>Major Risk</td></tr>
<tr><td rowspan="3">3.</td><td rowspan="3">B/Ds may have different technical readiness and capability in performing data sharing with CDC.</td><td colspan="3">Technical readiness and capability for data sharing (e.g. data captured in electronic or in paper forms, quality of data, etc.) may vary across the databases of different B/Ds. While the early stage of CDC development could be kick-started with databases with higher technical readiness and capability for data sharing, those with lower technical readiness and capability may delay future development/ enhancement of CDC if the technical readiness is not sufficient for data sharing with CDC. Hence, the likelihood for this risk is considered high with a medium impact.</td><td rowspan="3">• CDC Project office to provide technical support to B/Ds with a lower level of technical readiness and capability to ensure compatibility for data sharing
• Regular checkpoint on participating B/Ds databases to ensure alignment of technical readiness and capability</td></tr>
<tr><td>Likelihood</td><td>Impact</td><td>Priority</td></tr>
<tr><td>High</td><td>Medium</td><td>Major Risk</td></tr>
</table>

Annex I - Anticipated Relations and Contributions of CDC Platform Upon Data Alignment



Data Sources

DH-CASIS
Date of birth
Gender

DH-CHSS
Date of birth
Gender

HA-CMS / clinical database
Date of birth
Gender
District of residence

EDB-KGESS
Date of birth
Gender

EDB-SEMIS
Date of birth
Ethnicity
Gender

SWD-CRSRC
Age
Gender
District of residence

SWD-CPR
Age
Ethnicity
Gender
In HK since birth
New arrival

DH-SMASH
Age
Date of birth
Gender
Place of birth

HKPF-CMIS
Date of birth
Gender
Place of birth

EDB-STIMS
Date of birth
Ethnicity
Gender
District of residence

HA-CanSQS
Gender

DH-SMASH
ADHD, EBD, ID

DH-CASIS
ADHD, ASD, HI, ID, MI, MotorI, SpLD, SI, VD&CI

HA-CMS / clinical database *
ADHD, ASD, EBD, HI, ID, MI, MotorI, PD, SpLD, SI, VD&CI, VI

HA-CanSQS
VD&CII

EDB-SEMIS
ADHD, ASD, HI, ID, MI, PD, SpLD, SI, VI

SWD-CPR
ADHD, ASD, EBD, HI, ID, MI, PD, SpLD, SI, VD&CI, VI

DH-CASIS
Diagnosis information

DH-SMASH
Borderline intelligence
Referral action

EDB-SEMIS
Assessment information

SWD-CRSRC
ID, PD, VI

SWD-CPR
Details at school
Disability information

SWD-CPR
Housing type, Family structure
CSSA, Residential status
Other family conditions

DH-CHSS
Height, Weight
Education attainment of parents
Head circumference
Date of birth of parents
Background of mother
Other family conditions

HA-CMS / clinical database
Hospitalisation information

DH-SMASH
Height, Weight
Education attainment of parents
Housing type, Details at school
Education level
Other child conditions
Occupation of parents

EDB-SEMIS
Language at home
Kindergarten level

EDB-STIMS
Language at home
Details at school

SWD-CRSRC
Housing type
Family structure
Last class attended
Other child conditions
New arrival
Other family conditions

HA-CanSQS
Category of age at diagnosis
Number of subjects
Year of diagnosis

EDB-KGESS
Details at school
Kindergarten level

CDC Platform

Demographic data (B/Ds' systems)
Age
Date of birth
Ethnicity
Gender
District of residence

Demographic data (one B/D system)
New arrival
Place of birth
In HK since birth

Subject-specific data - SEN (B/Ds' systems)
Attention-deficit/hyperactivity disorder (ADHD)
Autism Spectrum Disorder (ASD)
Emotional and Behavioural Difficulties (EBD)
Hearing impairment (HI)
Specific Learning Difficulties (SpLD)
Visceral disability and chronic illness (VD&CI)
Mental illness (MI)
Motor impairment (MotorI)
Physical disability (PD)
Intellectual disability (ID)
Speech impairment (SI)
Visual impairment (VI)

Other SEN data (one B/D system)
Assessment information
Borderline intelligence
Details at school
Diagnosis information
Disability information
Reasons for referral service
Referral action

Other child-related data (B/Ds' systems)
Height, Weight
Education attainment of parents
Housing type
Language at home

Other child-related data (one B/D system)
School related (details at school, education level, kindergarten level, last class attended and other child conditions)
Health related (category of age at diagnosis, head circumference, number of subjects, year of diagnosis and hospitalisation information)
Family related (CSSA, family structure, new arrival, residential status, date of birth of parents, occupation of parents, background of mother and other family conditions)

CDC Platform Contributions

Demographic Data
To align the five demographic data across B/Ds

Incidence Rates
To find out the number of children with SEN by 12 types

Characteristics of the Children with SEN
To explore the characteristics of the children with SEN, and the profiles of parents

Reporting Gaps
To figure out the reporting gaps across databases and explore the underlying reasons

Risk Factors
To examine the risk factors including school-related, health-related and family-related factors of the children with SEN

Comparison of Children with SEN and those without SEN
To compare the characteristics and risk factors between children with SEN and those without SEN

Geographical Distribution
To illustrate the geographical distribution for children with SEN

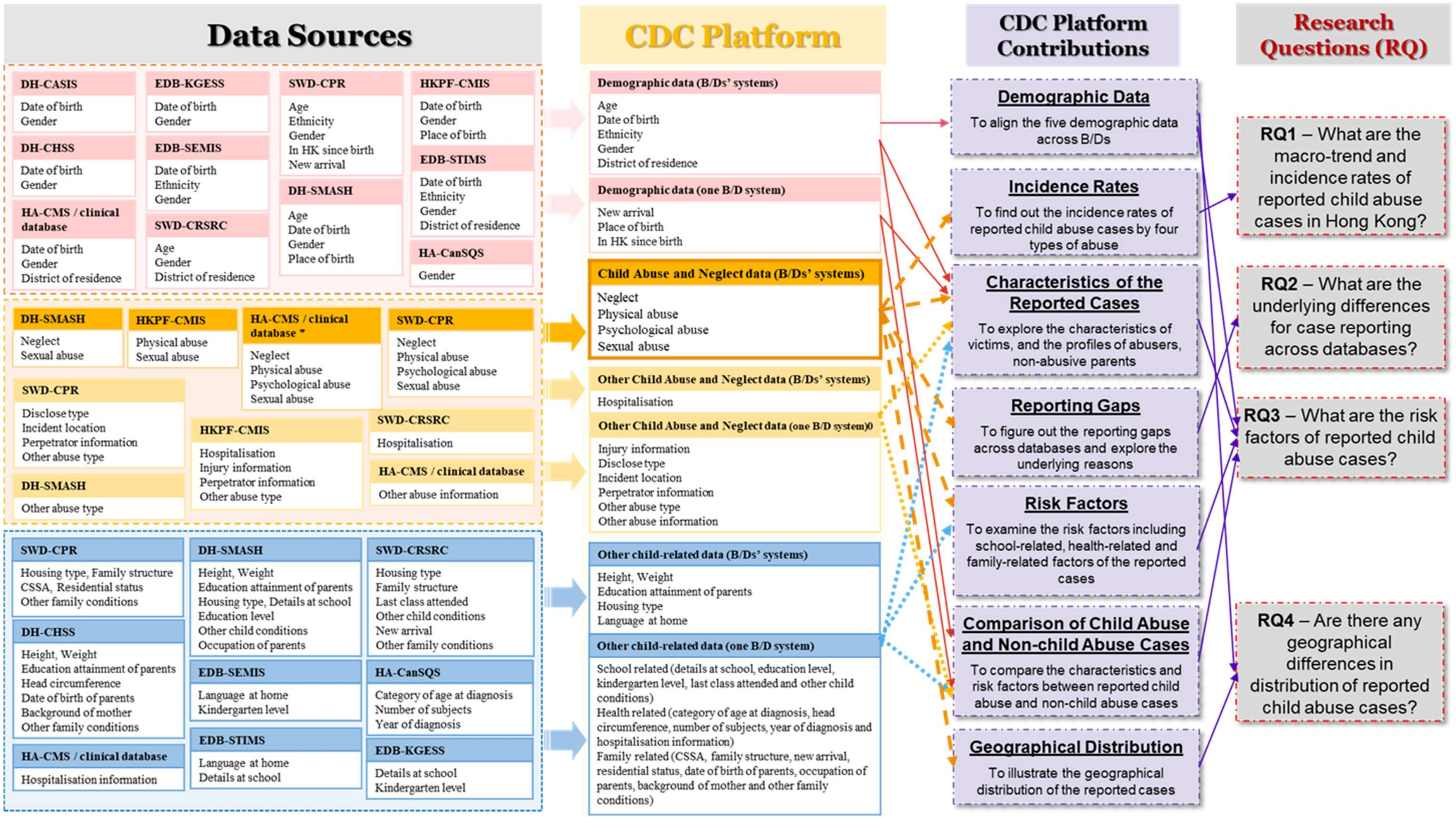
Research Questions (RQ)

RQ1 – What is the number of children with SEN in Hong Kong?

RQ2 – What are the factors contributing to differences in reporting of children with SEN conditions across various databases?

RQ3 – What are the characteristics of children with SEN that can influence health, safety and/ or education outcomes?

RQ4 – What is the prevalent mode and utilisation of health, educational and social service provision (i.e. across or within DC/ CA) for children with SEN?



Data Sources

DH-CASIS
Date of birth
Gender

DH-CHSS
Date of birth
Gender

HA-CMS / clinical database
Date of birth
Gender
District of residence

EDB-KGESS
Date of birth
Gender

EDB-SEMIS
Date of birth
Ethnicity
Gender

SWD-CRSRC
Age
Gender
District of residence

SWD-CPR
Age
Ethnicity
Gender
In HK since birth
New arrival

DH-SMASH
Age
Date of birth
Gender
Place of birth

HKPF-CMIS
Date of birth
Gender
Place of birth

EDB-STIMS
Date of birth
Ethnicity
Gender
District of residence

HA-CanSQS
Gender

DH-SMASH
Neglect
Sexual abuse

HKPF-CMIS
Physical abuse
Sexual abuse

HA-CMS / clinical database *
Neglect
Physical abuse
Psychological abuse
Sexual abuse

SWD-CPR
Neglect
Physical abuse
Psychological abuse
Sexual abuse

SWD-CPR
Disclose type
Incident location
Perpetrator information
Other abuse type

DH-SMASH
Other abuse type

HKPF-CMIS
Hospitalisation
Injury information
Perpetrator information
Other abuse type

SWD-CRSRC
Hospitalisation

HA-CMS / clinical database
Other abuse information

SWD-CPR
Housing type, Family structure
CSSA, Residential status
Other family conditions

DH-CHSS
Height, Weight
Education attainment of parents
Head circumference
Date of birth of parents
Background of mother
Other family conditions

HA-CMS / clinical database
Hospitalisation information

DH-SMASH
Height, Weight
Education attainment of parents
Housing type, Details at school
Education level
Other child conditions
Occupation of parents

EDB-SEMIS
Language at home
Kindergarten level

EDB-STIMS
Language at home
Details at school

SWD-CRSRC
Housing type
Family structure
Last class attended
Other child conditions
New arrival
Other family conditions

HA-CanSQS
Category of age at diagnosis
Number of subjects
Year of diagnosis

EDB-KGESS
Details at school
Kindergarten level

CDC Platform

Demographic data (B/Ds' systems)
Age
Date of birth
Ethnicity
Gender
District of residence

Demographic data (one B/D system)
New arrival
Place of birth
In HK since birth

Child Abuse and Neglect data (B/Ds' systems)
Neglect
Physical abuse
Psychological abuse
Sexual abuse

Other Child Abuse and Neglect data (B/Ds' systems)
Hospitalisation

Other Child Abuse and Neglect data (one B/D system)0
Injury information
Disclose type
Incident location
Perpetrator information
Other abuse type
Other abuse information

Other child-related data (B/Ds' systems)
Height, Weight
Education attainment of parents
Housing type
Language at home

Other child-related data (one B/D system)
School related (details at school, education level, kindergarten level, last class attended and other child conditions)
Health related (category of age at diagnosis, head circumference, number of subjects, year of diagnosis and hospitalisation information)
Family related (CSSA, family structure, new arrival, residential status, date of birth of parents, occupation of parents, background of mother and other family conditions)

CDC Platform Contributions

Demographic Data
To align the five demographic data across B/Ds

Incidence Rates
To find out the incidence rates of reported child abuse cases by four types of abuse

Characteristics of the Reported Cases
To explore the characteristics of victims, and the profiles of abusers, non-abusive parents

Reporting Gaps
To figure out the reporting gaps across databases and explore the underlying reasons

Risk Factors
To examine the risk factors including school-related, health-related and family-related factors of the reported cases

Comparison of Child Abuse and Non-child Abuse Cases
To compare the characteristics and risk factors between reported child abuse and non-child abuse cases

Geographical Distribution
To illustrate the geographical distribution of the reported cases

Research Questions (RQ)

RQ1 – What are the macro-trend and incidence rates of reported child abuse cases in Hong Kong?

RQ2 – What are the underlying differences for case reporting across databases?

RQ3 – What are the risk factors of reported child abuse cases?

RQ4 – Are there any geographical differences in distribution of reported child abuse cases?

Annex J - Pre-requisites to be fulfilled for implementation of Enhancement Mode of CDC

Dimension	Pre-requisites to be fulfilled for implementation of Enhancement Mode of CDC
Business and Legal Dimension	• Approval obtained from the oversight body of CDC with regard to the objective and scope of CDC's Enhancement Mode • Consultation with the Privacy Commissioner for Personal Data and engagement with experts (i.e. with expertises in data linkage and de-identification) • Establishment of a fully commissioned data ethics and privacy panel to ensure ethical use of children's data
Data Dimension	• Update of consent form of data providers to ensure future collection of data covers usage of de-identified data for data linkage projects • Data alignment to be performed for topics of concern based on consensus reached with data providers for linkage projects • Informed consent to be sought from data subjects for Enhancement Mode of CDC for data linkage projects and longitudinal studies where applicable
Technology Dimension	• A technically stable platform in place with well-defined operations and procedures
Implementation Dimension	• Satisfactory completion of PIA • Availability of funding for contracting out data linkage and/or longitudinal studies

Annex K - Indicative Stakeholder Engagement Plan

	Children's workforce, NGO service providers, academia and experts, and key stakeholders from the general public (e.g. parents/ guardians and children)	**Wider community**
Pre-implementation		
Objectives	• To brief purpose and upcoming involvement required for the implementation of CDC	• To brief purpose and potential benefits for the implementation of CDC
Key discussion topics	• Background of CDC development • Role of stakeholder group and upcoming support required	• Background of CDC development • Potential benefits for CDC
Mode(s) of engagement	• Briefing sessions	• Online and offline advertisement (e.g. physical leaflets, official videos and official publications on websites etc.)
Frequency	• One-off	• One-off
During implementation		
Objectives	• To gather business requirements for the implementation of CDC • To understand technical readiness of stakeholder group in data sharing • To identify potential challenges on data sharing	• To provide regular updates on CDC development
Key discussion topics	• Functionalities required in CDC design • Interfacing/ data sharing requirements • Current technical readiness and required support	• Status of CDC development • Anticipated next steps of CDC development

	• Potential challenges in data alignment, data processing and data privacy considerations	
Mode(s) of engagement	• Workshops • One to one discussion session, if required	• Official publications on Government websites
Frequency	• One-off (multiple rounds during system analysis and development phase of CDC implementation) • One to one discussion session, if and when required	• Quarterly/ bi-annual
Post-implementation		
Objectives	• To solicit feedback upon completion of each development milestone of CDC	• To solicit feedback from the wider community on the implementation of CDC and the pilot projects
Key discussion topics	• User feedback on CDC usage and adoption • Feedback on demonstrated benefits of pilot projects • Suggestions on topics for further research projects	• Feedback on pilot projects and introduction of CDC
Mode(s) of engagement	• Workshops • One to one discussion session, if required • Online communication channels	• Online communication channels • Community-led engagement
Frequency	• One-off for workshops (multiple rounds after each development milestone of CDC) • One to one discussion session, if and when required • On-going for online communication channels	• On-going