

**For discussion  
on 11 October 2021**

**Legislative Council Panel on Welfare Services**  
**Consultancy Study on**  
**Developing a Central Databank on Children**

**Purpose**

This paper briefs Members on the progress of a consultancy study on developing a Central Databank on Children (CDC) in Hong Kong.

**Background**

2. The Government has engaged a consultant to assist the Commission on Children (CoC) in conducting a study to explore the feasibility and implementation framework of developing a CDC in Hong Kong to enable the Government and/or relevant non-government stakeholders to collect and share useful data on children.

**Literature Review of Overseas Experience**

3. With reference to five overseas CDC projects as summarised in **Annex A**, the consultant has identified eight areas for consideration when developing a CDC in Hong Kong, with emphasis on “Development Objectives”<sup>1</sup>, “Scope/Types of Data Collection”<sup>2</sup> and “Application”<sup>3</sup>. A summary of the eight areas is at **Annex B**. The consultant has advised that 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 “Scope/Types of Data Collection”, which will in turn incur very different levels of data privacy

---

<sup>1</sup> “Development Objectives” include “trend monitoring” (covered in 4 out of 5 of the projects studied), “policy appraisal” (3 out of 5), “prevention & early intervention” (1 out of 5), and “case tracking” (1 out of 5).

<sup>2</sup> “Scope/Types of Data Collection” ranges from “health & safety”, “education & skills”, “material well-being”, “family & peer relationship” to “behaviour & risk”. All the 5 projects studied cover the first three dimensions, but only some projects cover the last two.

<sup>3</sup> “Application” includes “Enable collaboration with stakeholders across different sectors (included in all 5 projects), “Inform strategy formulation and policy evaluation” (4 out of 5), “Raise awareness and facilitate discussion” (4 out of 5), “Create a holistic narrative on children’s well-being” (4 out of 5), “Inform differential situations between different segments of children” (3 out of 5) and “Enable decision making on service provision” (1 out of 5).

concerns. For “trend monitoring”, collection of aggregate and anonymised data by leveraging on existing operating databases should suffice; whereas “case tracking” will require data which are identifiable and traceable.

## **Local Stakeholder Engagement**

4. The consultant has taken stock of 18 local databases from Government Bureaux/Departments (B/Ds) and six local databases from non-governmental organisations (NGOs)/academia. It is observed that local databases are primarily designed for “trend monitoring” and “prevention & early intervention” through the means of causal analysis and risk identification as well as for research purposes, covering the five major well-being dimensions under the “Scope/Types of Data Collection”. Majority of Government owned and NGO/academia owned databases cover data on children in general, with a fraction of them also covering data on specific segments of children. When compared with NGO/academia owned databases, Government owned databases have higher technical maturity (in terms of data governance and quality) and more advanced technical architectures. Government owned databases are more ready to support data sharing with the CDC, as compared with those owned by NGOs/academia, if a CDC is set up in the future. Those owned by NGOs/academia are usually set up in a one-off, time-limited manner, and the data therein are not updated continually.

5. To gauge the views, possible concerns and expectations of stakeholders, the consultant completed a stakeholder engagement exercise from January to May 2021, in which it interviewed relevant B/Ds, NGOs and academia that collect children related information; held focus groups involving organisations in various fields, including social welfare, parent support, schools and educators, social science and healthcare; conducted engagement sessions with the public; and completed a survey with a sample size of about 1 000 respondents. The consultant’s preliminary views on an initial set of Guiding Principles for the development of a CDC in Hong Kong are set out in paragraph 6 below.

## **Guiding Principles**

6. Having regard to overseas practices and views collated through the stakeholder engagement exercise, the consultant has proposed a set of guiding principles for CDC development :

- (a) ***Purpose-specific/policy-driven:*** CDC development should be purpose-specific with clearly defined objectives with due consideration to be given to its potential in contributing to the attainment of the strategic goals of the CoC.
- (b) ***Privacy protected and secured:*** Data privacy and system security of the CDC should be accorded the utmost importance when considering the type of databank and data governance mechanism for CDC development, taking full

account of 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:*** It is important for the CDC to incorporate features that are valued by users (e.g. data discovery, data visualisation and upload/download functions etc.), especially if collaboration with organisations outside the Government is required in terms of inputs of data on specific segments of the children population.
- (d) ***Collaborative, Transparent and Consultative:*** In order to build public trust for CDC development, the process of developing the CDC should be transparent, with 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 Government B/Ds. As such, it will likely span across a number of years as demonstrated by observed overseas experience. It is therefore suggested that the CDC should be developed in an incremental, scalable manner, including tasks and projects that can demonstrate its benefits.
- (f) ***Relevant and adaptable:*** Development of the 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:*** It is important to demonstrate the net value of a CDC by assessing its development costs against its benefits in terms of the capability in facilitating the Government's formulation/ modification of children-related policies and programmes on evidence-based reviews, collation of children-related data for compiling statistics for monitoring trends and/or comparison with relevant benchmarks on children well-being in overseas jurisdictions, tangible support for specific segments of children such as children with special educational needs (SEN) and children having the experience of abuse.

## **Key Parameters**

7. The CoC has adopted the following key parameters for the further study of developing a CDC in Hong Kong in the final stage :

- (a) ***Development objectives:*** Based on the consultant's analysis with reference to overseas experience, the "Development Objectives" of "trend monitoring" and "policy appraisal" on the basis of non-identifiable data for the general

children population should be adopted, while the feasibility of adopting “prevention & early intervention” for specific segment(s) of the children population should be further examined. In view of the concerns over data privacy expressed by stakeholders in the engagement exercise, “case tracking” for the general children population will not be pursued.

- (b) **Data governance:** In view of the current legislative framework for the protection of the personal data<sup>4</sup>, a special methodology will be developed to safeguard the personal data of the children concerned. Specifically, individual B/Ds will update their datasets containing personal data, and convert them into an anonymised dataset by applying a pre-defined set of algorithm (which will be kept confidential) for conversion of personal data (i.e. Hong Kong Identity Card/birth certificate numbers) therein into unique reference numbers (URNs). The CDC data administrator will then construct a combined dataset by matching the records in the anonymised datasets from different B/Ds using the URNs, and use the combined dataset for “policy appraisal” and/or “trend monitoring” according to the themes or topics to be determined by CoC. The Privacy Commissioner for Personal Data will be consulted on the above methodology in due course.
- (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”, the CDC administrator will keep a central databank with non-identifiable data from B/Ds for the purpose of “trend monitoring” and “policy appraisal”; and at the “micro-level”, the relevant B/Ds will retain data with identifiable data in their respective databanks for the purposes of “prevention & early intervention” and “case tracking”. The CDC data administrator will set data standards to be endorsed by the CoC for unified definitions/ formats to be adopted by different B/Ds, and administer the procedure and protocol of data inputs, data retrievals and data disposals, including the specific timeframe for data retention and the specific types of data that can be shared.
- (d) **Incremental approach:** With reference to overseas experience, the CoC has agreed that an incremental approach should be adopted in developing a CDC in Hong Kong by rolling out an “Initial Central Databank” in selected themes/segments first. Taking into account the concerns expressed by stakeholders in the engagement exercise in six priority areas<sup>5</sup>, the CoC has selected “Risk of Abuse & Neglect” and “Children with SEN” as the priority areas for further examination by the consultant in its final analysis, on the

---

<sup>4</sup> Data Protection Principle 3 in Schedule 1 to the Personal Data (Privacy) Ordinance (Cap. 486) stipulates that personal data shall not, without the prescribed consent of the data subject, be used (including “be transferred” and “be disclosed”) for a new purpose unless any applicable exemptions under Part 8 of the Ordinance applies.

<sup>5</sup> The six priority areas include “Risk of Abuse & Neglect”, “Children with SEN”, “Chronic Health Conditions”, “Living in Poverty”, “Special Family Background” and “Ethnic Minorities”.

basis of the following objective criteria: (i) whether the target segment can be clearly defined, and data could be easily collected and retrieved for meaningful analysis and policy formulation; (ii) whether relevant data of the target segment are currently captured by multiple Government departments/parties; and (iii) whether the potential impacts will be significant and measurable.

### **Next Step**

8. Depending on the progress of the consultant in finalising the report on a feasible implementation framework and realistic timeline of developing a CDC, the CoC aims to complete the consultancy study in around mid-2022.

### **Advice Sought**

9. Members are invited to note the progress and give views on the consultancy study.

**Labour and Welfare Bureau**

**October 2021**

## Summary of Findings from Review of Overseas Practices

<b>Project</b>	<b>United Kingdom – Children’s Well-being Measures</b>	<b>United Kingdom – ContactPoint</b>	<b>Australia – Children’s Headline Indicators (CHI)</b>	<b>Western Australia – Well-being Monitoring Framework (WMF)</b>	<b>Alberta, Canada – Early Child Development Mapping Project (ECMap)</b>
<b>Key Role</b>	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 18 <sup>th</sup> 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.
<b>Duration</b>	11 years, since conceptualisation phase in 2009	5 years (terminated in 2010)	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

<b>Project</b>	<b>United Kingdom – Children’s Well-being Measures</b>	<b>United Kingdom – ContactPoint</b>	<b>Australia – Children’s Headline Indicators (CHI)</b>	<b>Western Australia – Well-being Monitoring Framework (WMF)</b>	<b>Alberta, Canada – Early Child Development Mapping Project (ECMap)</b>
					community engagement)
<b>Governance</b>	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.
<b>Legislation</b>	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 of children.	Children well-being legislation provided a facilitative environment for research.
<b>Type of data</b>	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.

<b>Project</b>	<b>United Kingdom – Children’s Well-being Measures</b>	<b>United Kingdom – ContactPoint</b>	<b>Australia – Children’s Headline Indicators (CHI)</b>	<b>Western Australia – Well-being Monitoring Framework (WMF)</b>	<b>Alberta, Canada – Early Child Development Mapping Project (ECMap)</b>
<b>Implementation consideration</b>	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.
<b>Achievement or outcomes</b>	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.



**Eight Areas for Consideration when  
Developing a Central Databank on Children**

<b>(a) Development Objectives</b>	<ul style="list-style-type: none"> <li>• Monitoring trends and appraisal of strategic priorities, policies and programmes could be potential initial development objectives of the CDC.</li> <li>• Engaging stakeholders within and outside of the Government to prioritise key focus areas could facilitate consensus building.</li> <li>• A set of guiding principles shall also be considered to help guide the process of aligning expectation on the potential role of the CDC and its mandate in Hong Kong in the final phase of this Study.</li> <li>• Predictive application in the form of prevention and early intervention could be a value proposition of the CDC in the longer term.</li> </ul>
<b>(b) Legislative Framework</b>	<ul style="list-style-type: none"> <li>• The Personal Data (Privacy) Ordinance (Cap. 486) offers foundational personal data protection in Hong Kong.</li> <li>• Additional safeguards (either through legislative or administrative means) will need to be considered if the CDC seeks to share data of higher sensitivity.</li> </ul>
<b>(c) Institutional Governance</b>	<ul style="list-style-type: none"> <li>• Assignment of roles for undertaking oversight and executive functions for the CDC should be context-specific, giving due consideration to the objectives of the CDC, existing governance arrangement and public perception of the arrangement.</li> <li>• Composition of the governance body should consider the objectives of the CDC and be made up of critical parties.</li> <li>• Type of skills required will depend on the objectives and functions of the CDC, with the need for the assigned entity to consider capability building over time.</li> <li>• Setting up engagement platform will not only facilitate gathering of stakeholders' views, but also build trust amongst stakeholders in the process of CDC development.</li> </ul>

	<ul style="list-style-type: none"> <li>• Need for data governance body to make decisions on sharing of identifiable and traceable data for research purposes.</li> </ul>
<b>(d) Scope/Types of Data Collection</b>	<ul style="list-style-type: none"> <li>• Types of data to be collected should be fit-for-purpose in meeting the intended objectives and uses of the CDC.</li> <li>• There is a need to practically consider the availability of existing data with the need to address privacy, data ethics and other concerns of potential data contributors and data subjects.</li> <li>• Mechanism to enhance availability of data over time should be considered once the foundational/basic level of data has been incorporated into the CDC to meet its intended objectives.</li> <li>• Potential for considering more robust analysis through data linkage projects particularly for organisations with stored identifiers of their data subjects.</li> </ul>
<b>(e) Technical Infrastructure</b>	<ul style="list-style-type: none"> <li>• A user-centric design of the CDC with features that best meet the needs of potential users is likely to improve the chance of success for CDC implementation.</li> <li>• Development of the technical architecture should consider various components such as data access, storage and sharing as well as the extent to which existing IT infrastructure within the Government could be leveraged.</li> <li>• Potential to offer advanced data analytical functions in the longer term giving due consideration to the eventual objective and developmental stage of the CDC in the future.</li> </ul>
<b>(f) Data Management and Governance</b>	<ul style="list-style-type: none"> <li>• General alignment with overseas practices on measures adopted for safeguarding data security and privacy, with the continuous need to ensure that future safeguards in place are proportionate to the potential risks and sensitivity level of data collected/integrated in the CDC.</li> <li>• Organisations with lower level of technical maturity in terms of data quality and data operations may require support if their data are deemed to be critical for sharing through the CDC.</li> </ul>

	<ul style="list-style-type: none"> <li>• It is essential to consider a robust ethical data framework in relation to the collection and use of children-related data regardless of whether more sensitive data will eventually be made available through the CDC.</li> <li>• The “Five Safes”<sup>1</sup> model can be considered to assure the public that a stringent, consistent and transparent process has been put in place to safeguard disclosure of information of higher sensitivity for research purposes.</li> <li>• Heightened security measures must be considered if data that identify or are able to trace individuals would be disclosed.</li> <li>• Gaining public acceptability is an absolute pre-requisite when sharing identifiable and/or traceable data due to likely widespread public interest and scrutiny.</li> </ul>
<p><b>(g) Implementation Approach</b></p>	<ul style="list-style-type: none"> <li>• A phased implementation approach can be adopted giving due consideration to contextual differences and the applicability of common implementation components (including stakeholder engagement) for developing and operating the CDC.</li> <li>• Provided that data privacy and security concerns of the public can be addressed, the use of pilots for demonstrating benefits of the CDC could potentially increase public confidence on CDC development before wider execution.</li> <li>• There is potential to implement an economic cost benefit analysis to demonstrate the projected impact to society in the longer term.</li> </ul>

---

<sup>1</sup> The “Five Safes” model is commonly adopted to ensure rigorous arrangements to protect data confidentiality. The model consists of five key components, including:

- Safe People – Only authorised researchers can access the personal information
- Safe Project – Only authorised research that is feasible, lawful and provides a public benefit will be permitted
- Safe Setting – Conducted in a secured environment
- Safe Data – The access arrangements for the data will be commensurate with the risk of data disclosure
- Safe Outputs – Thorough checks of research outputs and code are in place to protect data confidentiality

	<ul style="list-style-type: none"> <li>• Proactive engagement of the community can be considered to build and sustain the momentum of CDC development over time.</li> </ul>
<b>(h) Application</b>	<ul style="list-style-type: none"> <li>• With due consideration to the expected benefits of key stakeholders, CDC development should provide applications that bring about such benefits while bearing the aligned objectives of the CDC in mind.</li> <li>• Wider benefits can be brought about by data linkage projects focussing on prevention and early intervention if privacy, security and ethical concerns can be addressed.</li> </ul>