OPEN DATA SURVEY

Prepared by

Vamoeurn Nimol

Júlia Garcia Puig

Consultants

USAID FROM THE AMERICAN PEOPLE

fhi360 THE SCIENCE OF IMPROVING LIVES

Internews Local voices. Global change.

អង្គការទិន្នន័យអំពីការអភិវឌ្ឍ Open Development Cambodia

តើទិន្នន័យបើកចំហជាអ្វី?

វាជាទិន្នន័យដែលបង្កើតឡើងដោយស្ថាប័នរដ្ឋ វិស័យឯកជន អ្នកស្រាវជ្រាវ ឬប្រជាពលរដ្ឋដែលគ្រប់គ្នាអាចយកមកប្រើប្រាស់បានក្នុងគោលបំណងអ្វីក៏ដោយ ដោយគ្មានការហាមឃាត់។

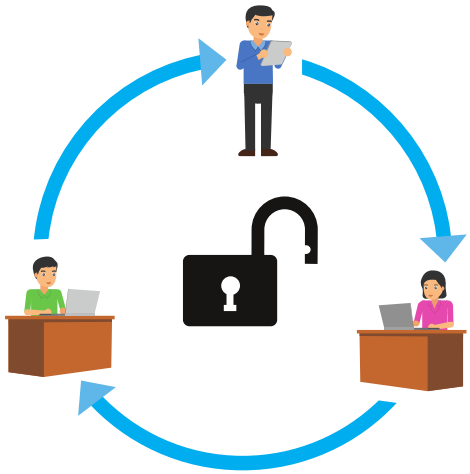
អ្នកសារព័ត៌មានជាអ្នកជ្រើសរើសថាព័ត៌មានអ្វីសំខាន់ដែលត្រូវចែករំលែក។

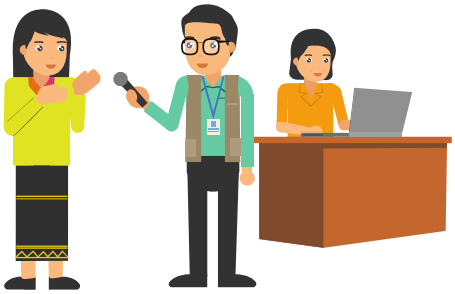
ប្រជាពលរដ្ឋចូលរួមក្នុងការផ្តល់យោបល់លើគោលនយោបាយ ហើយទាមទារគណនេយ្យភាពពីស្ថាប័នរដ្ឋ និងវិស័យឯកជន។

ការអភ្ជុតរំពីទិន្នន័យបើកចំហ

មានអ្នកផលិតនិងប្រើប្រាស់ទិន្នន័យ
ជាច្រើន។

“ប្រព័ន្ធទ្រទ្រង់” ទិន្នន័យបើកចំហដែលមាន
ដំណើរការល្អត្រូវការ “ការផ្គត់ផ្គង់” និង “តម្រូវការ”
ព្រមទាំងហេដ្ឋារចនាសម្ព័ន្ធមូលដ្ឋានសមស្រប។





ក្រុមការងារសិក្សារឿងទិន្នន័យបើកចំហ
បានពិនិត្យមើលទិន្នន័យដែលត្រូវបាន
ផលិតនៅក្នុងប្រទេសកម្ពុជានិងរបៀប
ដែលវាត្រូវបានប្រើប្រាស់។

ប្រទេសកម្ពុជាមានតម្រូវការខ្លាំង។
មានតែ ២០ ភាគរយនៃអ្នកឆ្លើយតប
ប៉ុណ្ណោះដែលផលិតទិន្នន័យ។ មាន
មនុស្សតិចណាស់ដែលចែករំលែក
ទិន្នន័យ។

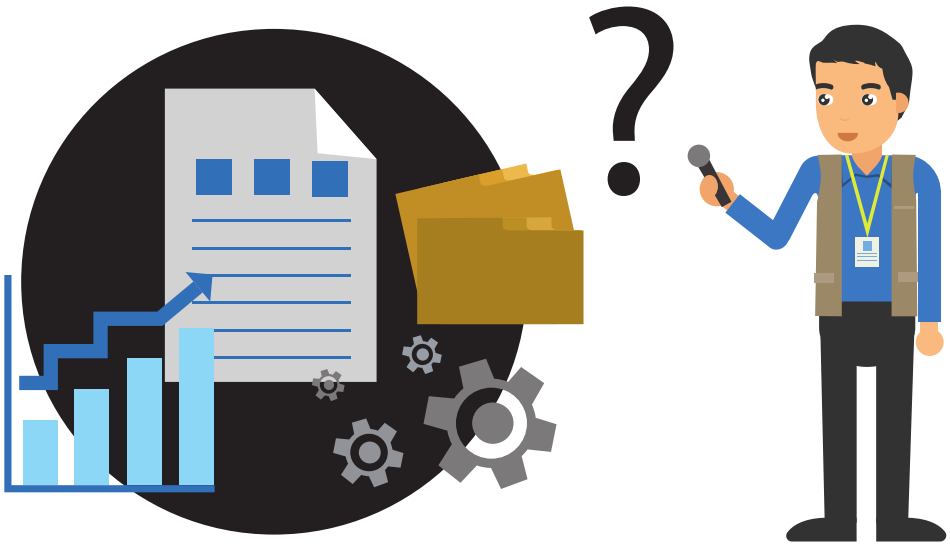
សាកលវិទ្យាល័យនានាបានប្រមូលទិន្នន័យ ប៉ុន្តែមានតិចតួចប៉ុណ្ណោះដែលចែកចាយវា។

អង្គការសង្គមស៊ីវិលផលិតទិន្នន័យភាគច្រើនសម្រាប់តែការប្រើប្រាស់ផ្ទៃក្នុងរបស់គម្រោង។

ប្រភពរដ្ឋាភិបាលច្រើនតែផ្តល់ជាព័ត៌មាន ប៉ុន្តែពុំសូវផ្តល់ជាទិន្នន័យដើមនោះទេ។

រដ្ឋាភិបាលពុំបានធ្វើមជ្ឈការទិន្នន័យច្រើននោះទេ។

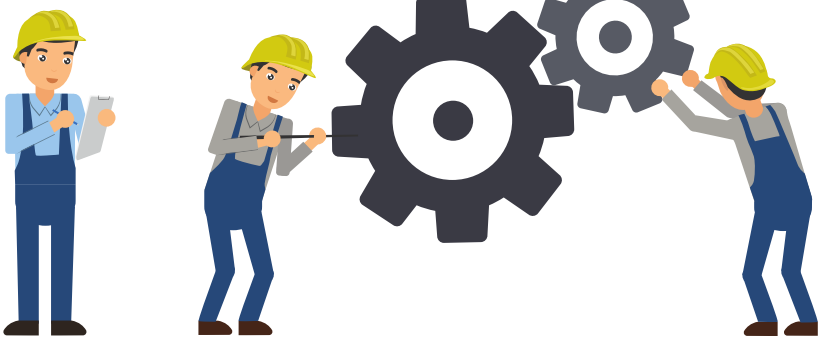
ទិន្នន័យបើកចំហនៅកម្ពុជាត្រូវបានប្រើប្រាស់ជាចម្បងសម្រាប់ការសិក្សាស្រាវជ្រាវ ការធ្វើរបាយការណ៍ និងការអនុវត្តគម្រោង។

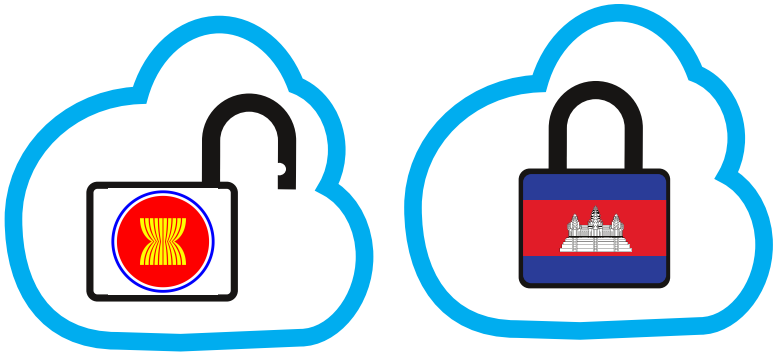


អ្នកសារព័ត៌មានភាគច្រើនមិនសូវយល់ដឹងពីការប្រើប្រាស់ទិន្នន័យសម្រាប់សរសេរព័ត៌មាននោះទេ។

មានអ្នកឆ្លើយសំណួរតិចតួចណាស់ដែលយល់ដឹងពីអ្វីដែលជា “ទិន្នន័យបើកចំហ” និង “អត្តទិន្នន័យ”

វត្តមានទិន្នន័យបើកចំហច្រើនបន្ថែមទៀតនឹងផ្តល់នូវការវាស់វែងកាន់តែច្បាស់អំពីសមិទ្ធិផលរបស់ប្រទេសកម្ពុជាលើបញ្ហាអភិវឌ្ឍន៍។





ប្រទេសជាច្រើននៅអាស៊ានកំពុងផ្តល់ការយកចិត្តទុកដាក់ខ្លាំងរឿងទិន្នន័យបើកចំហ និងបង្កើនការវិនិយោគលើហេដ្ឋារចនាសម្ព័ន្ធទិន្នន័យរបស់ពួកគេ។ តើកម្ពុជានឹងអនុវត្តដូចប្រទេសទាំងនោះដែរឬទេ? គម្រោងដែលនឹងត្រូវអនុវត្តមួយចំនួនហាក់ដូចជាផ្តល់ក្តីសង្ឃឹមខ្លះៗលើរឿងនេះ។

What is Open Data?

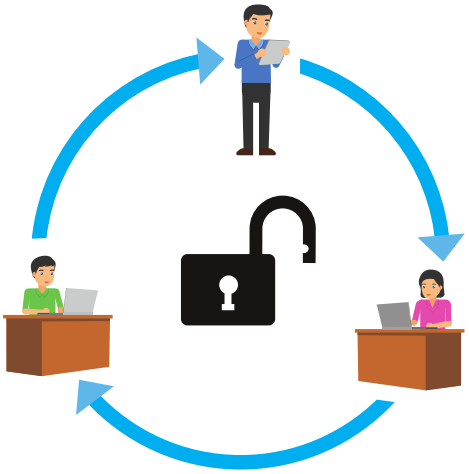
It is data generated by governments, business, researchers, citizens, for use by anyone - for any purpose, without restrictions.

Journalists identify key points to share.

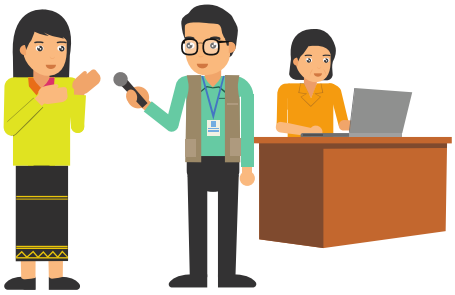
Citizens take action to inform policy and hold government and business accountable.

Open Data Survey

There are many producers and users of data.



A healthy open data 'ecosystem' has a 'supply' and 'demand', as well as an underlying infrastructure.



The Open Data Survey team reviewed data that is produced in Cambodia, and how it is used. This included an extensive review of websites and many interviews.

Cambodia has strong demand; only 20% of survey respondents produce data. Few actively share.

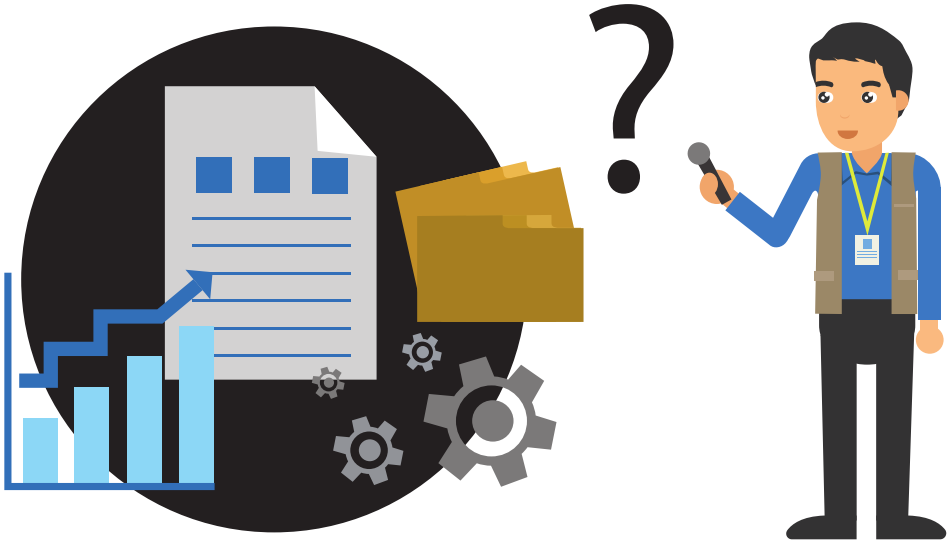
UNIVERSITIES collect data, but few produce/share it.

CIVIL SOCIETY produces data largely for internal project use.

GOVERNMENT sources provide information but not raw data.

And there is no strong centralization.

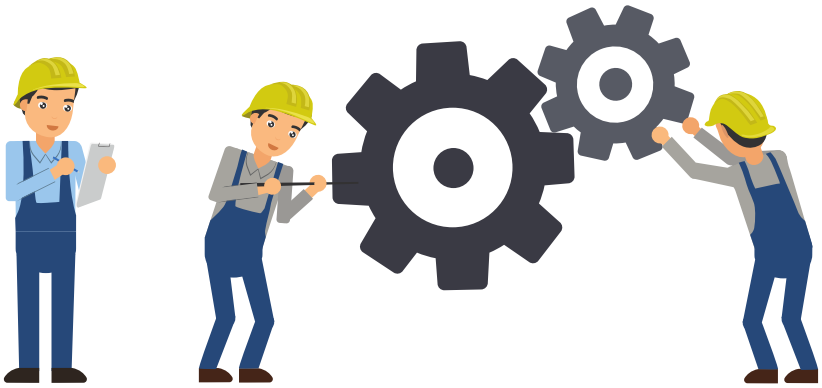
Main uses of open data in Cambodia are research, reports and projects.

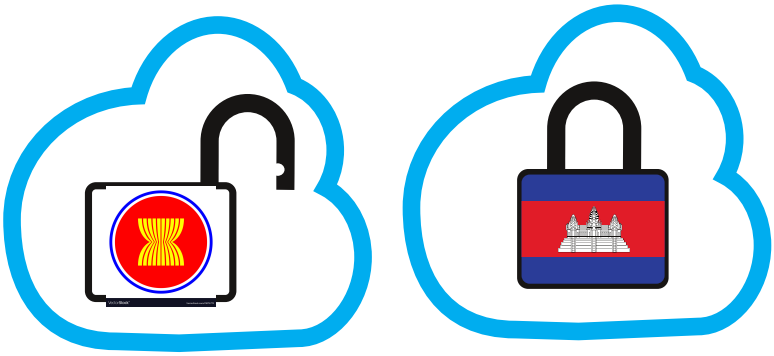


Journalists are largely not familiar with the use of data for stories.

Few respondents understand what 'open data' or 'metadata' is. Respondees note information gaps, unreliable data and poor metadata as key challenges.

More Open Data would provide clearer metrics for Cambodia's work on development challenges.





Many countries in ASEAN are prioritizing Open Data and investing in their infrastructure.
Will Cambodia join them? Some upcoming projects present some promise.

INNOVATIONS FOR SOCIAL ACCOUNTABILITY IN CAMBODIA

OPEN DATA SURVEY

Prepared by
Vamoeurn Nimol
Júlia Garcia Puig
Consultants

Submitted: May 13, 2020

Disclaimer

The findings, interpretations and conclusions expressed in this report are solely those of the authors and do not necessarily reflect the official views, policies, or positions of Internews, its staff or its affiliate entities. Assumptions made in the analysis are not reflective of the position of any individual or entity other than the authors. Since the content of this work involves critical analyses and thinking, the findings, interpretations and conclusions presented herein are always subject to change, revision and reconsideration in the light of new data or information which was not available to the authors during the course of carrying out this work. The content of this report, therefore, should not be held in perpetuity.

Authorship

This survey report was authored by:

Mr. Vamoeurn Nimol

✉ nimolva@gmail.com

(+855) 012 831 009 or 015 755 255

Vamoeurn Nimol has been a lecturer at the Institute of Technology of Cambodia (ITC) since 2005. When not involved in teaching, he works as a freelance consultant providing services in the spatial information technology field. He holds a master's degree in Spatial Information Technology from the University of Melbourne and a Master of Engineering in Environmental Geology from Gadjah Mada University. His expertise ranges from spatial programming and databases, spatial statistics and analysis, web mapping, spatial data infrastructures and the integration of ICT in community development.

Ms. Júlia Garcia Puig

✉ julia@opendevcam.net

(+855) 069 486 531

Júlia Garcia Puig is working as the Research and Strategy Officer for Open Development Cambodia (ODC). She has a background in economics and holds a master's degree in International Development from the Barcelona Institute of International Studies (IBEI). In her current role, Júlia contributes to ODC's strategic plan, conducts research, and produces publications on the Sustainable Development Goals, public service delivery and accountability. She also collaborates with the Ministry of Planning to advance the implementation of the Cambodian Sustainable Development Goals.

Table of Content

List of Tables

List of Figures

សេចក្តីសង្ខេប

របាយការណ៍អង្កេតនេះមានគោលបំណងស្វែងយល់អំពីតថភាពបច្ចុប្បន្ននៃធនធានទិន្នន័យបើកចំហនានាដែលមានស្រាប់ឬដែលកំពុងបង្កើតថ្មី ដែលពាក់ព័ន្ធនឹងដំណើរការគ្រប់គ្រងថវិការបស់គណៈកម្មាធិការជាតិសម្រាប់ការអភិវឌ្ឍតាមបែបប្រជាធិបតេយ្យនៅថ្នាក់ក្រោមជាតិ និងអំពីឱកាសនានាដើម្បីអភិវឌ្ឍវិស័យទិន្នន័យបើកចំហនៅកម្ពុជា។ ការអង្កេតនេះផ្តោតសំខាន់លើការកំណត់រកធនធានទិន្នន័យបើកចំហដែលមានប្រភពពីស្ថាប័នរដ្ឋាភិបាល សង្គមស៊ីវិល និងសាកលវិទ្យាល័យនានា។ ការអង្កេតនេះក៏បានចងក្រងចេញជាបញ្ជីប្រភពទិន្នន័យបើកចំហ និងវដ្តនៃការចេញផ្សាយទិន្នន័យនៅក្នុងប្រទេសកម្ពុជាផងដែរ។

ការអង្កេតនេះត្រូវបានធ្វើឡើងដោយផ្អែកលើការយល់ឃើញថាប្រព័ន្ធទិន្នន័យបើកចំហដែលមានដំណើរការល្អទាមទារឲ្យមានតុល្យភាពរវាងការផ្គត់ផ្គង់ (អ្នកផលិតទិន្នន័យ) និងតម្រូវការ (អ្នកប្រើប្រាស់ទិន្នន័យ) ដែលតុល្យភាពនេះអាចរក្សាបានដរាបណាហេដ្ឋារចនាសម្ព័ន្ធទិន្នន័យមានភាពរឹងមាំ។ ចំពោះផ្នែកតម្រូវការ ក្រុមអ្នកអង្កេតបានវាយតម្លៃលើភាពដែលអាចរកបាន និងភាពដែលអាចប្រើការបាននៃទិន្នន័យ របៀបគ្រប់គ្រងលើការប្រើប្រាស់ទិន្នន័យ និងការឲ្យតម្លៃទិន្នន័យបើកចំហពីសំណាក់អ្នកផលិតទិន្នន័យនានា។ ចំពោះផ្នែកផ្គត់ផ្គង់ យើងសិក្សាអំពីរបៀបដែលទិន្នន័យត្រូវបានប្រើប្រាស់ អំពីកង្វះខាតទិន្នន័យ និងអំពីរបៀបដែលអ្នកប្រើប្រាស់ឲ្យតម្លៃ និងចូលរួមចំណែកដល់ការអភិវឌ្ឍទិន្នន័យបើកចំហ។ សម្រាប់ហេដ្ឋារចនាសម្ព័ន្ធទិន្នន័យវិញ ក្រុមការងារបានសិក្សាលម្អិតអំពីមុខងារ សមត្ថភាព និងភាពងាយស្រួលប្រើនៃផតថលវ៉ិប (web portal) ឬគេហទំព័រដែលផ្តល់លទ្ធភាពឲ្យអ្នកប្រើប្រាស់អាចទាញយកទិន្នន័យបាន។

តាមរយៈការសិក្សាឯកសារដែលមានស្រាប់ ក្រុមការងារបានវាយតម្លៃគេហទំព័រច្រើនជាង ១០០ ដែលត្រូវបានចាត់ទុកថាជាទីតាំងសក្តានុពលដែលអ្នកប្រើប្រាស់អាចទាញយកទិន្នន័យបើកចំហបាន។ ចំពោះការវាយតម្លៃលើការផ្គត់ផ្គង់និងតម្រូវការ យើងបានសម្ភាសអ្នកផ្តល់ព័ត៌មានសំខាន់ៗចំនួន ២៣ នាក់ ដែលរួមមានសាស្ត្រាចារ្យឬអ្នកស្រាវជ្រាវមកពីសាកលវិទ្យាល័យនានាចំនួន ៨ នាក់ បុគ្គលិកសង្គមស៊ីវិល ៤ នាក់ អ្នកសារព័ត៌មាន ៤ នាក់ អ្នកប្រឹក្សាយោបល់ឯករាជ្យ ៣ នាក់ និងមន្ត្រីរដ្ឋាភិបាល ៤ នាក់។ ការសម្ភាសទាំងនេះត្រូវបានបំពេញបន្ថែមដោយការស្ទង់មតិសាធារណៈមួយដោយប្រើប្រាស់កម្រងសំនួរឌីជីថល ហើយដែលទទួលបានការឆ្លើយតបចំនួន ៧៩ នាក់។ ជាសរុប មនុស្សចំនួន ១០២ នាក់ត្រូវបានស្ទង់មតិ ដែលក្នុងនោះមានស្រី ២៦ នាក់។

ឆ្លងតាមការវិភាគនៃទិន្នន័យដោយផ្អែកលើសំណាកគំរូដែលប្រមូលបាន វិសាលគមទិន្នន័យនៅក្នុងប្រទេសកម្ពុជាមានទិសដៅបាញ់ឆ្ពោះទៅរកការប្រើប្រាស់ ជាជាងការផលិត។ ការផលិតទិន្នន័យមានតែចំនួន ២០ ភាគរយនៃអ្នកឆ្លើយតបប៉ុណ្ណោះ។ ការចែកចាយទិន្នន័យពុំសូវទទួលបានការយកចិត្តទុកដាក់នោះទេ ដោយមានអ្នកឆ្លើយសំណួរតែ ១៣ ភាគរយប៉ុណ្ណោះដែលបាននិយាយថាស្ថាប័នរបស់ពួកគេបានចែកចាយទិន្នន័យដែលពួកគេបានផលិត។ លើសពីនេះទៀត នៅមានការយល់ដឹងទាបពីភាពខុសគ្នារវាងទិន្នន័យ និងព័ត៌មានក្នុងចំណោមអ្នកប្រើប្រាស់និងអ្នកផលិតទិន្នន័យ។

ការដាក់ឲ្យប្រើប្រាស់ទិន្នន័យបើកចំហនៅតាមស្ថាប័នសិក្សាស្រាវជ្រាវនានា ដូចជាសាកលវិទ្យាល័យជា

ដើម នៅពុំទាន់មានសកម្មភាពនៅឡើយទេ ទោះបីជាអ្នកស្រាវជ្រាវបានប្រមូលទិន្នន័យយ៉ាងច្រើនក៏ដោយ។ មានមូលហេតុចំបងចំនួនបួន ដូចជាកម្មសិទ្ធិបញ្ញា ការយល់ថាទិន្នន័យនឹងគ្មានអត្ថប្រយោជន៍អ្វីក្នុងការប្រើប្រាស់ក្រៅពីគម្រោងស្រាវជ្រាវផ្ទាល់ ការខ្លាចការរំលោភលើល័ក្ខខ័ណ្ឌដែលអ្នកឧបត្ថម្ភការស្រាវជ្រាវបានដាក់ឲ្យ និងអវត្តមានវេទិកាចែករំលែកទិន្នន័យដែលអាចទុកចិត្តបាន ដែលបានចូលរួមចំណែកវារាំងពុំឲ្យមានការចែករំលែកទិន្នន័យដែលបានប្រមូលរួច។ ចំពោះអង្គការសង្គមស៊ីវិលវិញ ការផលិតទិន្នន័យតែងតែត្រូវបានបម្រុងទុកសម្រាប់ប្រើប្រាស់ក្នុងគម្រោងផ្ទាល់របស់ស្ថាប័ន ឬសម្រាប់ចែករំលែកដល់ក្រុមការងារតែប៉ុណ្ណោះ ហើយទិន្នន័យត្រូវបានគេចាត់ទុកថាគ្មានតម្លៃសាធារណៈ ឬផ្តុំកព័ត៌មានឯកជនច្រើនពេក។ ទោះយ៉ាងណាក៏ដោយ លទ្ធផលដែលបានពីទិន្នន័យទាំងនេះ ជារឿយៗត្រូវបានបោះពុម្ពផ្សាយនៅក្នុងរបាយការណ៍គម្រោងរបស់ពួកគេ។ នៅថ្នាក់ជាតិ ក្នុងចំណោមគេហទំព័ររដ្ឋាភិបាលជាង ៥០ ដែលក្រុមអ្នកអង្កេតបានពិនិត្យសិក្សានោះ មួយភាគធំផ្តោតលើការផ្តល់ព័ត៌មាន ហើយមិនផ្តោតលើការផ្តល់ទិន្នន័យដល់សាធារណៈជនឡើយ។ នៅពេលដែលទិន្នន័យត្រូវបានផ្តល់ឲ្យអ្នកប្រើប្រាស់ ទិន្នន័យស្ទើរតែទាំងស្រុងត្រូវបានបង្ហាញជាទម្រង់សង្ខេបរួម ដូចជាតារាងតួលេខ ជារូបភាព ឬ អត្ថបទនៅក្នុងឯកសារ PDF ដែលពេលខ្លះគ្រាន់តែជាកំណែឌីជីថលដែលបានពីការស្កេនសំណៅដើមប៉ុណ្ណោះ។ តាមពិតនៅគ្រប់លំដាប់ថ្នាក់ទាំងអស់ រដ្ឋាភិបាលកម្ពុជាពុំមានផតថលវ៊ិបសម្រាប់ចែករំលែកទិន្នន័យនោះទេ។ វាជាការណ៍ដែលគួរកត់ឲ្យសំគាល់ ដែលថាអង្គការសហប្រជាជាតិបានចែករំលែកទិន្នន័យអំពីប្រទេសកម្ពុជាដល់សាធារណៈជន ច្រើនជាងរដ្ឋាភិបាលកម្ពុជាខ្លួនឯង។

ទាក់ទងនឹងការប្រើប្រាស់ទិន្នន័យវិញ ស្ទើរតែ ៩០ ភាគរយនៃអ្នកឆ្លើយសំណួរបានបង្ហាញថាពួកគេបានប្រើប្រាស់ទិន្នន័យពីស្ថាប័នផ្សេង ដើម្បីបំពេញតម្រូវការទិន្នន័យរបស់ពួកគេ ដែលក្នុងនោះ ៨២ ភាគរយគឺដើម្បីគាំទ្រការស្រាវជ្រាវ ៧៤ ភាគរយសម្រាប់ការសរសេររបាយការណ៍ និង ៤៧ ភាគរយសម្រាប់ការអនុវត្តគម្រោង។ ទោះបីជាការវាយតម្លៃពេញលេញអំពីអក្ខរកម្មទិន្នន័យបើកចំហ និងអាកប្បកិរិយានៃការប្រើប្រាស់ទិន្នន័យមិនត្រូវបានធ្វើឡើងក៏ដោយ ក៏ភ័ស្តុតាងនានាមួយចំនួនបានបង្ហាញពីកម្រិតអក្ខរកម្មទិន្នន័យទាប និងឥរិយាបថលើការគ្រប់គ្រង និងការប្រើប្រាស់ទិន្នន័យបើកចំហដែលគួរឱ្យព្រួយបារម្ភ។ មានតែបីភាគរយប៉ុណ្ណោះនៃអ្នកចូលរួមក្នុងការអង្កេតនេះ ដែលបានយល់ដឹងត្រឹមត្រូវអំពីនិយមន័យនៃទិន្នន័យបើកចំហ។ ស្របគ្នានេះដែរ អ្នកសារព័ត៌មានទាំងអស់ដែលបានចូលរួមសម្ភាសបានឲ្យដឹងថាពួកគេមិនដឹងថាអ្វីដែលជាទិន្នន័យបើកចំហនោះទេ។ ពួកគេក៏បានអះអាងផងដែរថាសហសេវិករបស់ពួកគេនៅក្នុងវិស័យសារព័ត៌មានពុំត្រូវបានបណ្តុះបណ្តាលឱ្យធ្វើការជាមួយទិន្នន័យស្មុគស្មាញនោះទេ ហើយពួកគេខ្វះចំណេះដឹងលើការប្រើប្រាស់ទិន្នន័យ។ មួយភាគធំនៃអ្នកឆ្លើយសំណួរ (៤៥%) មិនដឹងថាអត្តទិន្នន័យ (metadata) មានន័យដូចម្តេចនោះទេ។

កង្វះខាតខាងហេដ្ឋារចនាសម្ព័ន្ធទិន្នន័យបើកចំហ និងអក្ខរកម្មទិន្នន័យបើកចំហទាប បានបង្កើតឲ្យមានការលំបាកយ៉ាងខ្លាំងសម្រាប់សហគមន៍ទិន្នន័យបើកចំហ។ ចន្លោះប្រហោងព័ត៌មាន ទិន្នន័យដែលមិនអាចជឿទុកចិត្តបានដែលបណ្តាលមកពីកង្វះខាតកំណត់ត្រាទិន្នន័យនិងអត្តទិន្នន័យ និងគុណភាពទិន្នន័យទាប គឺជាបញ្ហាប្រឈមធំបំផុតបីដែលអ្នកប្រើប្រាស់ទិន្នន័យបានឲ្យដឹង។ ទោះយ៉ាងណា បញ្ហាប្រឈមនៃការអភិវឌ្ឍពិតប្រាកដកើតឡើងពីកត្តាទាំងក្នុងនិងក្រៅ។ នៅពេលដែលប្រទេសជាច្រើននៅក្នុងប្លុកអាស៊ានកំពុងបើកឃ្លាំងទិន្នន័យរបស់ខ្លួនឲ្យសា

ធារណៈជនរបស់គេប្រើប្រាស់ និងកំពុងវិនិយោគយ៉ាងច្រើនទៅលើហេដ្ឋារចនាសម្ព័ន្ធទិន្នន័យ មានចម្ងល់ថាតើប្រជាជនកម្ពុជានឹងប្រឈមមុខជាមួយសេដ្ឋកិច្ចឌីជីថលដែលមានការផ្លាស់ប្តូរយ៉ាងឆាប់រហ័សបែបណា ហើយអាចរក្សាបានសមត្ថភាពប្រកួតប្រជែងយ៉ាងដូចម្តេច? កង្វល់ជាច្រើនសម្រាប់រដ្ឋាភិបាលកម្ពុជាជុំវិញបញ្ហាប្រឈមនានាដែលបណ្តាលមកពីកង្វះសមត្ថភាពទិន្នន័យទាំងនេះ ជាប់ពាក់ព័ន្ធសំខាន់ទៅនឹងការរៀបចំខ្លួនរបស់រដ្ឋាភិបាលក្នុងការវិនិយោគលើការផលិតទិន្នន័យថ្មីៗ និងការកសាងប្រព័ន្ធចាកចាយទិន្នន័យ ដែលអាចធានាឲ្យមានលំហូរទិន្នន័យជាប់លាប់ទៅដល់ដៃអ្នកប្រើប្រាស់ ដែលនឹងឆ្លើទិន្នន័យទាំងនោះឲ្យក្លាយទៅជាផលិតផលឬសេវាកម្មប្លែកៗដែលមានលក្ខណៈឆ្លើប្រឌិតខ្ពស់។

បញ្ហាដែលបណ្តាលមកពីទិន្នន័យធំបំផុតដែលជះឥទ្ធិពលដល់គ្រប់វិស័យនៃការអភិវឌ្ឍគឺការមានទិន្នន័យបើកចំហតិចតួចបានកាត់បន្ថយសមត្ថភាពឃ្លាំមើលនិងតាមដានរបស់សាធារណៈជនទៅលើគំនិតផ្តួចផ្តើមអភិវឌ្ឍន៍នានារបស់រដ្ឋាភិបាល។ ការត្រួតពិនិត្យលើរដ្ឋាភិបាលតាមរយៈការប្រើប្រាស់ទិន្នន័យពិបាកនឹងសម្រេចបាន លើកលែងតែក្នុងករណីដូចជា គជអប ដែលបានផ្តល់ឲ្យសាធារណៈជនបានដឹងពីគំនិតផ្តួចផ្តើមអភិវឌ្ឍន៍នានារបស់រដ្ឋាភិបាលតាមរយៈមូលដ្ឋានទិន្នន័យសាធារណៈចំនួន ៩ របស់ពួកគេ។

ថ្វីបើមានបញ្ហាប្រឈមជាច្រើនក៏ដោយ វានៅមានឱកាសមួយចំនួនសម្រាប់ការលូតលាស់នៃទិន្នន័យបើកចំហ។ ឧទាហរណ៍មួយគឺគម្រោងពង្រឹងសមត្ថភាពស្ថាប័នឧត្តមសិក្សា (HEICP) ដែលគាំទ្រថវិកាដោយធនាគារពិភពលោក ដែលក្រោមគម្រោងនេះ ការស្រាវជ្រាវវិទ្យាសាស្ត្រជាច្រើននឹងត្រូវបានគាំទ្រ។ គម្រោងទាំងនេះនឹងត្រូវការទិន្នន័យបើកចំហច្រើនដោយជៀសមិនរួច ហើយទីបំផុតនឹងជំរុញឱ្យមានកំណើនការផលិតទិន្នន័យជាច្រើនផងដែរ។ ឧទាហរណ៍មួយទៀតគឺរដ្ឋាភិបាលកម្ពុជាបានចាប់ផ្តើមទទួលស្គាល់ពីសារៈសំខាន់នៃផតថលទិន្នន័យបើកចំហដូចដែលបានបង្ហាញតាមរយៈការដាក់ឱ្យប្រើប្រាស់នូវផតថលទិន្នន័យរបស់ក្រសួងបរិស្ថាន។ ចំណាប់អារម្មណ៍មួយចំនួនលើប្រព័ន្ធព័ត៌មានភូមិសាស្ត្រស្ម័គ្រចិត្តក៏ត្រូវបានសង្កេតឃើញកើតមានផងដែរ។

Executive Summary

This survey report intends to formulate an up-to-date picture of available and emerging open data resources on the National Committee for Sub-national Democratic Development (NCDD) budgeting process and opportunities to grow the open data community in Cambodia. The survey focused on identifying open data resources made available by the government, civil society, and universities. The survey process also undertook a mapping of existing open data platforms and data release cycles in Cambodia.

This survey was carried out based on the understanding that a healthy open data ecosystem is built on a supply (data producers) and demand (data users) equilibrium that rests on robust data infrastructures. On the demand side, the survey team evaluated open data availability and accessibility, how it was governed and the appreciation of open data among data producers. On the supply side, it considers how the data is used, what data gaps are perceived, and how the users appreciated and contributed to the open data community. For the data infrastructures, the team looked exclusively at functionalities of data portals/websites and how user-friendly they were.

Through the desk study more than 100 websites – those perceived as the most likely locations where users can find open data – were reviewed. For the supply and demand assessment, we interviewed 23 key informants, consisting of eight university lecturers or researchers, four civil society workers, four journalists, three freelancers and four government officers. These interviews were complemented by a publicly administered web-based survey that resulted in 79 responses. In total 102 individuals were surveyed, among which 26 were female.

Based on the analysis of the collected data, Cambodia's data spectrum in our sample tilts significantly towards consumption. Production accounts for only 20 percent of the respondents. Data dissemination was even less common, with only 13 percent of the respondents stating that their organization disseminates data. There is also a low understanding of the differences between data and information among data users and producers.

Open data at research institutions such as universities is virtually nonexistent even though researchers collect huge amounts of data. Four main reasons – intellectual property, no usefulness outside of the project, fear of violation of terms and conditions of the sponsors, and no trustworthy sharing platform – contribute to collected data remaining private. For civil society organizations, data production is almost always intended for project use or closed group sharing and is considered as having no public value or containing too much private information to be shared. However, often the results of the research are published in their project reports. At the national level, among more than 50 government websites reviewed, a large portion is information-based and does not provide the public with any data. When data is provided, it is almost always presented in an aggregated form such as tables, figures, charts or in-text numbers in PDF documents, which are sometimes just the scanned digital versions of the original printout. The government of Cambodia practically has no centralized data portal at all levels of governance. It is noteworthy that the UN has made more data about Cambodia available to the public than the Cambodian government itself.

In terms of usage, almost 90% of the respondents revealed that they used data from sources outside of their work to meet their data consumption needs, of which 82% is to support research, 74% report writing, and 47% project implementation. Although comprehensive assessment of open data literacy and attitude was not conducted, anecdotal evidence indicates

a low data literacy level and worrisome attitudes towards open data stewardship and usage. Only three percent of those taking part in this study correctly defined what constitutes open data. In parallel, all journalists interviewed intimated that they did not know what constitutes open data. They also claimed that their colleagues in the press sector were not trained to work with complicated data and lack functional data literacy. A large proportion of the questionnaire respondents (45%) do not know what metadata means.

The deficiency of enabling open data infrastructures and low open data literacy, combine to create an incredibly challenging ecosystem for the open data community. Information gaps, unreliable data due to poor data documentation and metadata, and low data quality are the top three perceived challenges faced by the data users. Real development challenges, however, come from both internal and external factors. As many countries in the ASEAN bloc are opening up their data repositories to their public and investing heavily on enabling data infrastructures, one would wonder how the Cambodian people will deal with a fast-changing, data driven digital economy and remain competitive. The dilemma surrounding these data driven changes for the Cambodian government centers around how well-prepared it is to invest in data production and delivery systems that ensure a reliable flow of data to the hands of its people who will then turn it into innovative and creative products and services.

The biggest data-induced problem that crosscuts every development aspect is that the limited availability of open data undercuts the public's oversight of the government's development initiatives. Government oversight through data is hard to achieve, except for the NCDD which provides the public some snapshots of the government's development initiatives through their nine public databases.

Despite these many issues, some opportunities for the growth of open data exist. An example is the World Bank funded Higher Education Institutions Capacity Improvement Project (HEICP), under which many scientific research projects will be supported. These projects will inevitably require a lot of open data and eventually will drive huge data growth. Another example is that the Cambodian government has begun to appreciate the importance of open data portals, as demonstrated by the launch of the Ministry of Environment's data portal. Some interests in voluntary geographic information systems were also observed.

1 Background

The realization that innovation and creativity cannot thrive in silos of protectionism or in closed and disconnected environments is perhaps the main driving force of technological revolution in the last two decades. Openness and transparency are achievable through open data, and accompanying technologies, standards and policies can be viewed as significant enablers for social accountability. The logic behind this is simple. A more open and transparent development facilitates citizen engagement which results in greater accountability and larger contribution to responsible development. This engagement happens when the citizen is well-informed and knowledgeable about the development process initiated by the government. This knowledge is made possible only when the citizen has access to data used and produced by different institutions. With this knowledge comes demands for better services and enhanced accountability from the government for its development initiatives.

Access to public services information, for example, can serve to illustrate the importance of citizen engagement in the state's service delivery. Data such as the location of healthcare facilities, expertise of healthcare practitioners, and service ratings can help ordinary citizens to make critical decisions that can mean the difference between life and death. Sending a critically ill patient to the nearest hospital that does not have capable doctors might delay the treatment. Another example is the importance of public data in enhancing transparency and accountability in development projects. When the government decides to invest in building infrastructure (for example a highway) it ensures transparency and enforces accountability by releasing data on the bidding process, the winning bid, bid values, estimated time of completion, socio-economic and environmental impacts, actual expenditure and fines. When the public knows that the highway construction should last only six months, any delay will incur fines and possible legal responsibility. This knowledge allows the public to act as a monitor to ensure that public well-being is not sacrificed to please a slow contractor. When environmental impact data of a hydropower project is open, the public whose livelihood depends on the health of the ecosystem can rally support to stop the project or to hold the project owner accountable for fair compensations to affected communities.

Cambodia has rapidly transitioned toward a lower middle-income country during the past decade. It has managed to maintain a steady annual economic growth of around seven percent.[1] Despite this positive development, there remain some underlying institutional weaknesses – one of which is the lack of information on development efforts being rolled out or implemented by the government – that are potentially constraining Cambodia's inclusive and sustainable social, economic and environmental development. With the freedom of information law still in limbo[2] despite having been drafted for a few years and the recognition that access to data and information is crucial for balanced and responsible economic growth, there are pressing needs to assess how information is made available to the public.

To contribute to addressing this critical issue and promoting inclusive development, this report presents the results of the ISAC Cambodia Open Data Survey which took place between mid-January and April 2020. The survey is framed under the project *Innovations for Social Accountability in Cambodia* (ISAC) which is a five-year (2019-2024) project funded by the

[1] World Bank, Benefiting from the Digital Economy: Cambodia Policy Note, 2018. Accessed Jan. 21, 2020

[2] Phnom Penh Post, Review of information law to be completed by month's end, Feb. 14, 2020. Accessed Mar. 30, 2020

United States Agency for International Development (USAID) and implemented by FHI 360, with its partners Development Alternatives Incorporated (DAI), Open Development Cambodia and Internews. ISAC is implemented in parallel with the second phase of the Implementation Plan for Social Accountability (ISAF 2) project, a joint initiative implemented by Cambodia's National Committee for Sub-National Democratic Development (NCDD) and a host of Civil Society Organizations (CSOs). Currently, ISAC supports demand side social accountability activities in seven municipalities. The survey was carried out by a team of service contractors (authors of this report) on behalf of Internews, an ISAC partner and international nonprofit organization working on building healthy media and information environments to inform and empower citizens to make informed decisions, participate in their communities, and hold power to account.

2 Objective

Aiming to contribute to the overall understanding of the Cambodian open data landscape, this survey intends to formulate an up-to-date picture of available and emerging open data resources on the National Committee for Sub-national Democratic Development (NCDD) budgeting process and opportunities to grow the open data community. The survey will focus on identifying open data resources made available by the government, civil society, and universities. The survey process will also result in a data mapping of existing NCDD-related open data platforms and data release cycles in Cambodia.

3 Assessment Method

This survey was carried out based on two guiding principles. First, the survey posits that a healthy open data ecosystem is built on a supply and demand equilibrium. This means that both the data users and producers, which represent the demand and supply sides of the equation respectively, play an equally important role in maintaining the well-being of the system. Second, it was presumed that the whole ecosystem is carried by a wide range of infrastructures.

Based on these principles, to formulate an accurate picture of the open data situation in Cambodia, it is important to assess three different but intertwined elements of the ecosystems:

- **Supply-side assessment:** This involves a critical analysis of the data producers. The survey explored three important indicators. First, the availability and accessibility of the existing data were assessed. Second, we looked at how these resources were governed and how their uses were dictated. This involved the assessment on data governance policies and data licensing practices. Finally, a behavioral assessment on the producers' appreciation of open data was investigated. It is undeniable that the government is the biggest producer. However, CSOs, international development agencies, universities, research institutions and the UN bodies are also contributing significantly to the open data ecosystem.
- **Demand-side assessment:** In a similar fashion to the analysis of the producers, the survey assessed the data users based on three indicators. First, how the data was being used and perceived data gaps were explored. Next, the team assessed how the users contributed back to the open data community. The third indicator again measured the user's behavior in relation to the appreciation of the open data.

- **Infrastructure assessment:** Finally, the survey assessed the role of infrastructures in enabling the open data ecosystem. These infrastructures include data portals and websites being run by different data producers. Two indicators are used for this assessment. First, the survey looked at whether each data producer had a centralized data repository. Also assessed was how user-friendly the producer's data sharing platform is.

Additionally, to ensure the consistency and uniformity of the findings, this survey used a single definition (discussed in the following section 3.1) of what constitutes open data.

3.1 Defining Open Data

As the appreciation of open data increases,[3] and even more individuals and institutions sign up to it, it becomes ever more important that a clear and agreed definition of what constitutes open data is established if we are to fully reap the benefits of openness and avoid its potential technical and legal risks. Semantically, what makes data open rests entirely on how open is defined. The Open Knowledge Foundation (OKF) makes precise the meaning of "open" with respect to knowledge in its 2005 Open Definition.[4] Under this definition, open can be summarized as "being free for anyone to access, use, modify and share—subject, at most, to measures that preserve provenance and openness." There are two inherently important elements to data openness under this definition. First, open data must have technical openness which means technical barriers to using the data must not be present. In this sense, open data must be machine readable and available in bulk. As an example, data provided in a digital format from a scanned printout is extremely hard to work with and therefore is not technically open. Even data that is given as tables in PDF documents (which are hard to extract) do not qualify as open. The second aspect implies legal openness which dictates how people access, use, and share the data. To qualify for legal openness, the data must be governed by an appropriate license that allows anyone free access to the data and the ability to reuse, build on, share, or place it into the public domain.

This definition of "open" is built on three important principles that make open data so powerful. These principles are what we look for when assessing whether data qualifies as being open. The following is a short summary on the meaning and interpretation of the three principles addressed in the OKF's Open Data Handbook.[5]

- **Availability and Access:** This principle dictates that data must be available as a whole (preferably on the internet) in a convenient and modifiable format. This principle does not necessarily require that the data be available free of charge, but somehow necessitates that, should it not be free of charge, the data shall not incur unreasonable reproduction costs.
- **Reuse and redistribution:** This generally means that data producers are not allowed to place conditions on how people can use the data, but are permitted to require that users must always properly credit them, that changes to the data must be clearly specified, or that new datasets derived from their data be licensed as open data.

---

[3] Canares M., Young A. and Verhulst S., 'Open Development Cambodia: Opening information on development efforts' in *Open Data in Developing Economies: Toward building an evidence Base on what works and how*, ed. Verhulst S. and Young A. (African Minds, 2017) 98.

[4] "Open Definition 2.1." *The Open Definition.* Accessed Feb. 12, 2020. https://opendefinition.org/od/2.1/en/

[5] "What is Open Data" *Open Data Handbook.* Accessed Feb. 12, 2020. https://opendatahandbook.org/guide/en/what-is-open-data/

- **Universal participation:** This principle ensures that discrimination against fields of practice, persons or groups is strictly prohibited. This means restrictions such as “non-commercial use only” or “only for humanitarian purpose” are not allowed.

3.2 Approach

To ensure the success of this work, a fact-based, results-oriented, and inclusive assessment approach that aims to deliver outputs of the highest quality possible within the specified time frame was devised. It can be broken down into three components as discussed below:

- **Desk study:** The first step of the approach was to conduct a desk study to build a proper context of the overall open data landscape. The desk study focused on open data mapping and policy analysis. The survey team visited the websites or data portals of important data producers to assess the condition of the existing data availability and accessibility and supporting infrastructures. Existing data governing policies, strategies or plans were also reviewed. These fulfill some parts of the second and third indicator of the supply-side assessment respectively and the two indicators of the infra-structure assessment.
- **Key informant interviews:** To complement the desk study, we conducted face-to-face, phone or virtual interviews with key informants. They are university teachers, journalists, and data producers. This activity fulfills indicator three of the supply-side and demand-side assessments. It also contributes to all other indicators of the three assessment components described above.
- **Questionnaire-based survey:** A digital and self-administered, highly structured questionnaire was deployed to the public to collect data on all the indicators of the three assessment elements. It contains 24 bilingual questions (English and Khmer). The use of digital questionnaires increases data reliability and quality and reduces cost and time. The collected data is digital and ready for analysis without the need for data encoding.

3.3 Methodological Limitations

Although we strive to achieve the highest possible quality, this survey has some unavoidable and inherent methodological limitations, as discussed below:

1. **Infrastructure assessment:** This work exclusively involved assessing web pages. The survey team did not attempt to assess non-cyber infrastructures. This means physical data clearing houses or repositories were not investigated. The web page analysis was also limited to a page depth of four levels. This means relevant contents that were buried deeper in the page structure were not analyzed and thereby were considered as nonexistent. Furthermore, assessing the availability of open data on each website depended heavily on using the site’s provided search engine with four keywords: Cambodia, data, database, dataset. Ten search results were then selected randomly for data analysis.
2. **Data release cycle**: Two approaches are generally used to assess data release cycle. The simplest one uses the data release schedule published by each data producer as the release cycle. The second approach involves metadata analysis of recurrent datasets to estimate the release cycle, assuming that the metadata contains data release dates. Owing to the potentially large number of datasets and the fact that most datasets do not come with

proper metadata after all (based on the survey team's experience), metadata analysis was not considered in this study. The survey relied exclusively on published data cycles.

3. **Data reproduction cost**: The definition of open data allows data producers to charge a reasonable reproduction cost on the data they make available. In this study, the reproduction cost is eliminated from the definition. In this sense, only data available for download over the internet (which is considered as practically free of charge although is theoretically not) without any payment requirement is considered as open data.

4 Fieldwork Results

Field surveys took place from January 20 to February 10, 2020, during which five journalists, four CSO representatives, three freelance consultants, four government officers and eight university researchers were interviewed. Due to time constraints and scheduling arrangements, interviews with the four government officers were done over the phone. Interviews with three university researchers and two CSO representatives were conducted virtually using VoIP instant messaging. Although some informants gave us their approval to be identified, all interview records and notes were anonymized to protect the identity of the informants.

In addition to the interviews, the research team also ran a public digital survey using a web-hosted questionnaire. In total, 79 responses from various individuals were received. Adding this number to the total interviews returns a sample size of 102 observations. Table 1 provides a summary of the characteristics of the survey's sample.

Table 1. Type of informants interviewed for the study

Type	Female	Male	Total
University lecturer/researcher	3	5	8
CSO	2	2	4
Journalist	0	4	4
Freelancer	1	2	3
Government officer	1	3	4
Questionnaire respondents*	19	60	79
Total	26	76	102

*Respondents provided inputs by completing the questionnaire

In terms of educational separation, the majority of the informants had post-graduate degrees (49 MSc and 17 PhD degree holders). There were also 11 students who took part in the survey. In terms of gender distribution, 26 women and 76 men were interviewed for the study (last row of Table 1). Table 2 summarizes the educational differences within the sample.

Table 2: Level of education of the informants

Key informant	BSc	MSc	PhD	Total
University lecturer/researcher	0	3	5	8
CSO	1	3	0	4
Journalist	n.a	n.a	n.a	n.a
Freelancer	1	2	0	3
Government officer	3	1	0	4
Questionnaire respondents	25	40	11	76
Total	30	49	17	95

5 Analysis and Findings

The following section presents the findings obtained from the interviews, what was stated by the respondents of the questionnaire and the reviews of websites and policies.

5.1 Production and Dissemination of Data

The evaluation of the production, dissemination, and consumption of data among the study's participants is a synthesis of the results from key informant interviews and the responses to questions contained in the questionnaire. The findings (summarized in Figure 1) indicated that data production, dissemination and consumption are highly imbalanced. The data spectrum uncovered by this survey tilts significantly towards consumption. Production accounts for only 20 percent of the respondents. Data dissemination gets even lower priority, with only 13 percent of the respondents stating that their organization disseminates data.

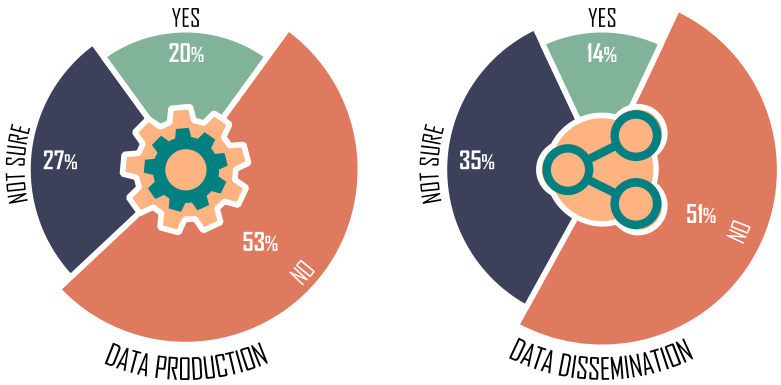


Figure 1: State of data production and dissemination (sample size: 93)

Based on the interviews with journalists, CSO representatives and government officers, the data production and data consumption spectrum can be linked to a low understanding of the differences between data and information. Journalists, in particular, stated that they did not know much about data-driven journalism. This problem exhibited itself more clearly when we looked at the data production at government institutions. Among more than 50 government websites reviewed, a large portion of them (refer to Table 5 and Table 8 in the Appendices) does not provide the public with any data. These are information-based websites where stories about the government's activities were featured on most of the pages. When data was indeed offered, the usability can be low. Data is almost always presented in an aggregated form such as tables, figures, charts, or in-text numbers in PDF documents, which are sometime the scanned digital versions of the original printout. Data collection methods were often not provided.

For university researchers, data creation is one of their core responsibilities. Through academic projects, researchers are constantly creating new primary data. Surprisingly, the survey's review of the websites of four major universities found no published data. The unavailability of data can be attributed to four main reasons. Research data is strictly considered by the interviewed researchers as their intellectual property whose scientific values remain long after it was produced. And because the aim of any research project is to get results published in either scientific journals or communities, unpublished data will never be shared publicly. Researchers do share their project data within their small circle of peers. Secondly, most of the researchers surveyed do not see their data as being useful outside of their designed purpose.

They believe no one can make use of the data due to its highly specialized creation methods and narrow purpose.

> *"You know what, as a researcher, I still find it exceedingly difficult to understand my research data. Who else can make use of it when they do not even have enough context?"* – A biotechnology researcher at the Institute of Technology of Cambodia.

Another inherent roadblock preventing the release of research data to the public is the terms and conditions of the research project sponsors. Unless clearly stated in the project's contract that data should be made public once it has been used for scientific purposes, the researchers will not release it for fear of violating the terms and conditions of the agreement or upsetting the project's backers. The survey team were able to briefly review three research project agreements (one being funded by JICA, another being funded by the Swedish government, and the third one being a joint project funded by a local NGO) and found no specific provision in the documents that dictates public data sharing. Finally, the researchers revealed that even though they are willing and allowed to share the data, they do not have any trustworthy data sharing platform to make their data available. When asked whether they knew that Open Development Cambodia has a data contribution mechanism, all the researchers interviewed stated that they were unaware of this feature.

For civil society organizations (CSOs), data production is almost always intended for internal use or closed group sharing. This data type was produced to support project implementation and is considered as having no public value. Two CSOs, both of which were involved in agricultural market research, were interviewed and both revealed that their data simply cannot be shared because it contains a lot of private information about its informants. When asked why they did not anonymize the data first before releasing it to remove any privacy concerns, both believed the process only adds an unnecessary burden to their already limited resources. These CSOs, however, reassured us that all their work is published in their project reports. However, in our understanding, project reports contain aggregated data, synthesized analysis results and opinions of their authors that make them highly subjective. Raw data, on the other hand, is more neutral and allows for many possible reuses and reinterpretations.

> *"Our data was mostly collected from face-to-face interviews in which names, addresses and contact numbers of the informants are aplenty. We can't release this privacy-sensitive data to the public. And should we want to remove privacy content from it, we simply don't have enough human resources to scan through all the datasets. You know, it cannot be automated. It must be processed by human eyes."* – An officer of an international CSO working on agricultural value chains.

5.2 Data Usage

The reason stated in the earlier section that the Cambodia's data spectrum tilts significantly towards consumption is based on the finding that a lot more people are using data to support their work. Almost 90% of the respondents reported that they used data from sources outside of their work to meet data consumption needs. Among these people, data is crucial for their research (82%), report writing (74%) and project implementation (47%). Figure 2 provides a summary of how data is being used in Cambodia.

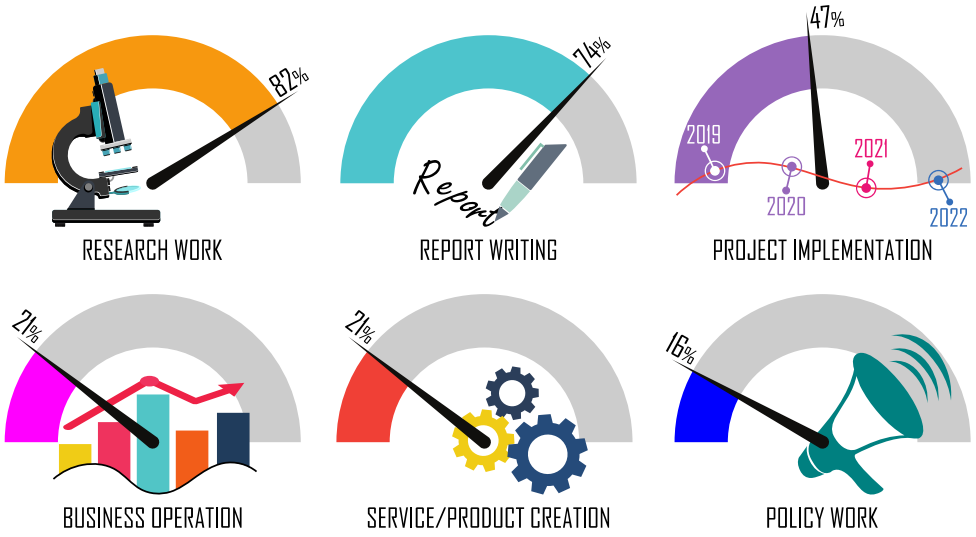


Figure 2: Purpose of data use (sample size: 57 respondents out of 95)

This raises a question about who supplies the data to meet all these consumption needs. Figure 3 illustrates that government agencies (including the National Institute of Statistics, NIS) and Open Development Cambodia are the two main sources of data. But this contradicts our findings regarding the low availability and accessibility of data on the government's websites.

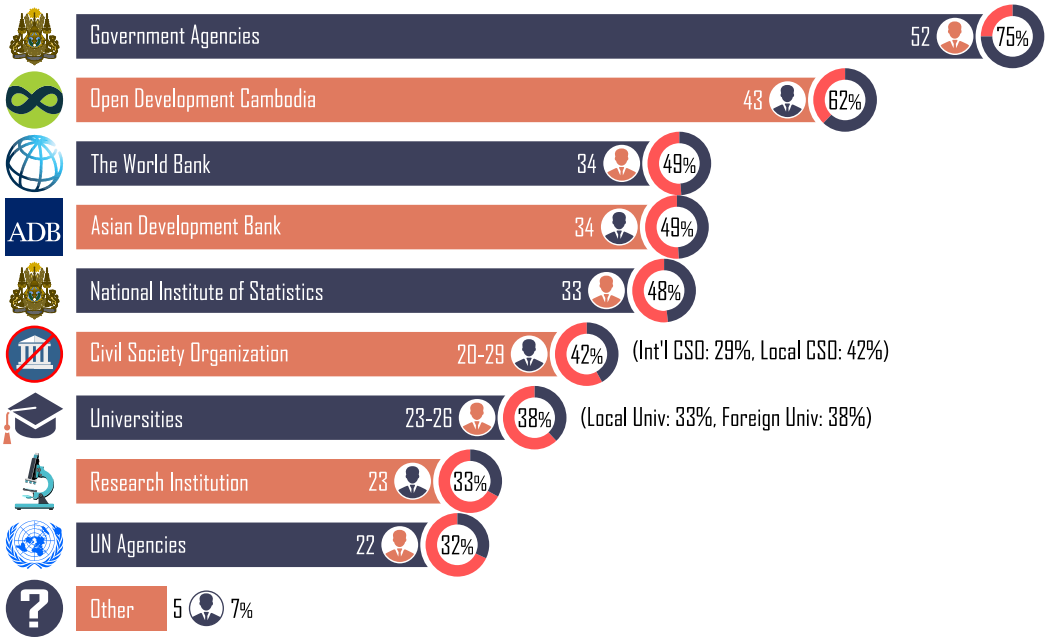


Figure 3: Data sources used by the survey respondents (sample size: 69 respondents out of 95)

The inability of the users to access the data through the government's websites and the likely dependence on informal access signal two potential data usage problems. The first one is related to whether unlawful payment for the data has been solicited. This concern manifests itself in Figure 4. Almost half of the respondents (47%) admitted that they paid external data producers for access to their data. Payments made to government institutions were reported by 31% of the respondents. Despite this, it is impossible to determine whether payment made to government institutions are unlawful because some agencies such as the National Institute of Statistics (NIS) are permitted to sell licensed data to the public.

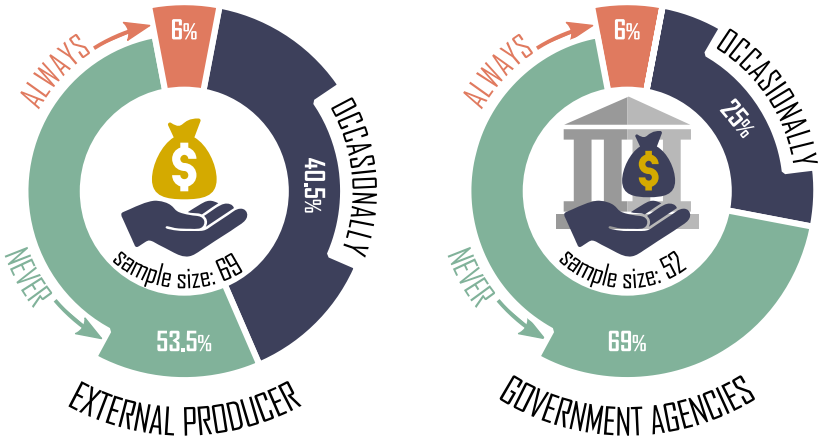


Figure 4: Payment to external data producers for data access

The second problem involves data integrity. Informal sharing of data is the biggest source of data integrity problems because changes made to the data cannot be tracked systematically and documented. Users will not have the ability to assess the quality of the data and are thereby faced with higher risks of information, misrepresentation, and misinterpretation.

When it comes to choosing data providers, data users view data reliability (62%) and the reputation of the data producers (48%) as the top two decision criteria (Figure 5).

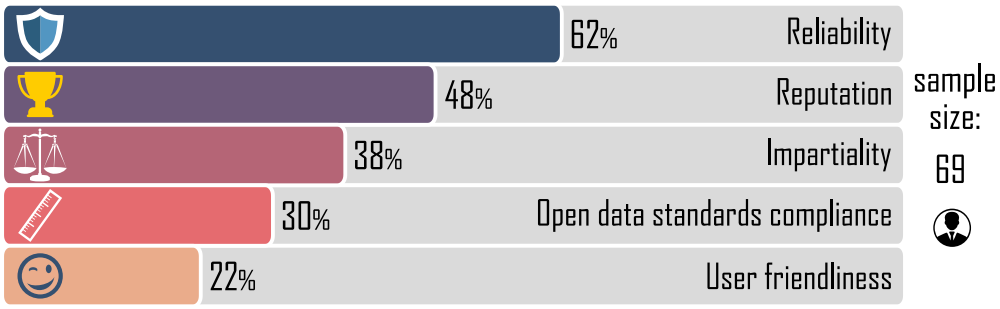


Figure 5: Criteria used for choosing data producers

5.3 NCDD and Its Data Usage

The review of the NCDD data portal revealed that nine databases are made available for public consumption (No. 5 of Table **6** in the Appendices). These include the Project Implementation Database, Gazetteer Database Online, Commune Database Online, Commune Council Database, NCDD Library, Bidding Announcement Database, ISAF M&E Database System, DMK M&E Internal System and Sub-national Project Database. The following section explains the data held within each database:

- Project Implementation Database (pID): This database was designed to record data related to the implementation of projects financed by the Commune/Sangkat Fund, which constitutes the budget allocated by the Cambodian government to the local administrations to support local development efforts. The main aspect of the pID database centers on data related to each implemented project. It includes project name and nature, development sector/sub-sector, awarded contractor, identity of bidders and their corresponding bid value, project start and end dates, project supervisors and project outputs/progress.

- Cambodia Gazetteer (gZ): This database stores the official names of all villages, communes/sangkats, districts/khans and provinces of Cambodia. The names are recorded in Khmer and in Romanized forms and are identifiable by a unique code.
- Commune Database Online (cDB): This database is perhaps the largest of all. It stores demographic data of each commune/sangkat in Cambodia. This includes general information such as population size (grouped by age and sex), number of households, occupations, earning and employment, housing status; economic information such as infrastructure, services, agriculture, industry; social information such as education/schooling, health status, water and sanitation, vulnerable groups; natural resources and environment information such as forest-dependent households, access to garbage collection, use of chemical fertilizers and pesticides; administration and security information such as birth registration, migration status, crimes, land disputes; gender information such as domestic violence, trafficking, women's migration; and information on minority groups.
- Commune Council Database (cCD): This database stores data related to each elected commune councilor. This includes name, sex, date of birth, Prakas, wealth declaration date, and political party.
- Bidding Announcement Database (BID): This database complements the Project Implementation Database (pID) discussed above. It is a very simple database that holds data related to the bidding process of each project such as bidding venue, application submission place, date of announcement, bidding placement time, bidding deadline, time for the opening of bidding applications and other project-related details.
- Implementation of the Social Accountability Framework (ISAF) Database: The database contains data on four areas: district, commune, primary school, and health center. Under each area, data related to work performance (including staff and infrastructure), revenue and expenditure can be found. Under the district and commune areas, two additional data types are given. Information related to district or commune projects and the Joint Accountability Action Plan (JAAP) can also be found.
- District/Municipality/Khan Monitoring and Evaluation System (DMK M&E): This database contains data on nine M&E areas, including 1) warrants/decisions/guidelines, 2) activity/budget implementation, 3) indicators, 4) revenue/expenditure, 5) cash assets, 6) inventory, 7) incoming documents/letters, 8) outgoing documents/letters, and 9) staff.
- Sub-National Project Database (sPD): While the pID database stores project implementation data at commune level, the sPD contains data on development projects up to the provincial level.

In terms of currency, the Commune Database Online and the Commune Council Database are significantly outdated with the former being last updated in 2010 and the latter in 2013. The results from our interview with NCDD reveal that although the publicly accessible Commune Database Online is out of date, the Commune Database itself is constantly updated and the data is available offline. The data migration from the offline database to the web has been hampered by the lack of technical and human resources since 2012. NCDD is currently trying to use its technical team to continue the data migration to make the Commune Database Online up to date. Despite this, NCDD assured us that organizational users can request any specific offline data if they submit an official request letter. It is unclear whether the public will also be provided the same access.

For the Commune Council Database, the transfer of data custodianship to a new

department is the cause for the delay in data updates. The new department does not have enough trained officers to manage the data update. The NCDD's secretariat is cooperating with the new department on technical training to allow for the resumption of the database's update. Given the importance of the two databases for ISAC implementation, currency issues can greatly affect their potential usefulness.

The currency of the other databases, on the other hand, is very recent. Based on NCDD's assertion, they were last updated as recently as 2019 or early 2020. However, based on our actual database evaluation, the currency of the data can vary from one province to another. It is worth remembering that the updating of the databases is highly dynamic. Except for the ISAF Database, the Sub-national Project Database and the Commune Database which are updated every year (a release cycle of one year), all other databases are updated irregularly depending on the occurrences of new data or projects. The Commune Council Database is a special case. It gets updated when new councilors have been elected. Although these databases are made available to the public, their uses are largely unknown. We are unable to ascertain its usage penetration. NCDD does not record the number of visits to its databases by the public, except for access to its library that averages around 20 per month. This means usage statistics are not available to evaluate the public's interest in the data. However, based on our interviews with key informants and a publicly administered questionnaire, less than 40 percent of the respondents use the NCDD's data (left chart of Figure 6). The majority of the actual users access the data through NCDD's website (right chart of Figure 6). The data is used mostly to support project implementation (85%). The use of data for research comes in second at 77%.

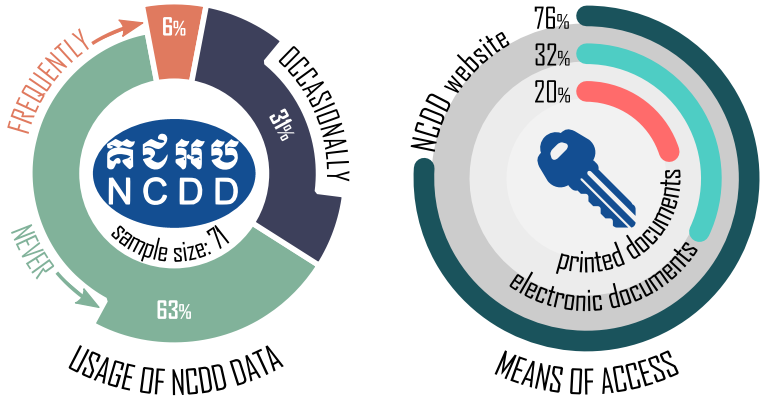


Figure 6: Usage of and means of access to NCDD's data

It is also particularly important to note that our study could not evaluate the accuracy, collection methods or completeness of the data released by the NCDD. We had no capacity and time to do such an assessment. The results from the questionnaire and key informant interviews also did not return any substantial conclusions. The data in the databases cannot be downloaded in bulk, either, which is very typical for databases. Instead, the users can submit queries for specific data and extract it individually.

5.4 Open Data Infrastructures

The production and consumption of the data discussed in the previous sections relies on another important factor. The two processes cannot exist without relevant and functional bearing infrastructures. If we look at the definition of open data discussed in section 3.1 (p. 3), data becomes realistically open only when it is accessible and downloadable through the internet.

This implies that cyber-infrastructures are the prerequisite for realizing open data endeavors. In a general sense, this means if a producer places its data on its website for public consumption, the data becomes open. In the strictest and most practical point of view, this general sense is not enough. Data that resides too deep in a web page's structure is less discoverable by search engines, making it almost unavailable for typical users. And placing a large amount of data on a website without a clear organizational structure and search mechanism is not making data discovery any easier. In a more practical sense, open data is useful only when it is easily discoverable by the users. Discoverability is what makes data portals so crucial as the enabling infrastructure of the open data ecosystem.

Recognizing data portals as an especially important infrastructure, this study reviewed more than one hundred websites perceived as the most likely sources where open data can be found. Our findings are wide-ranging. The Government of Cambodia practically has no centralized data portal at all levels of governance. Although data is made available at some institutions' websites–such as at the sites of the National Bank of Cambodia, the National Committee for Disaster Management and the National Institute of Statistics (row 3, 4 and 6 of Table 6 in the Appendices)–it is scattered across different sites and servers. This arrangement makes it hard for the users to find what they need unless they spend a huge amount of time scavenging for data on more than 50 government websites. This issue is reflected in our interviews with the university researchers and four journalists. All of them acknowledged that when they need data, their first approach is to use Google Search because they simply do not know any single point-of-access to the data. Other than this, they went to ODC's website to find the data they needed.

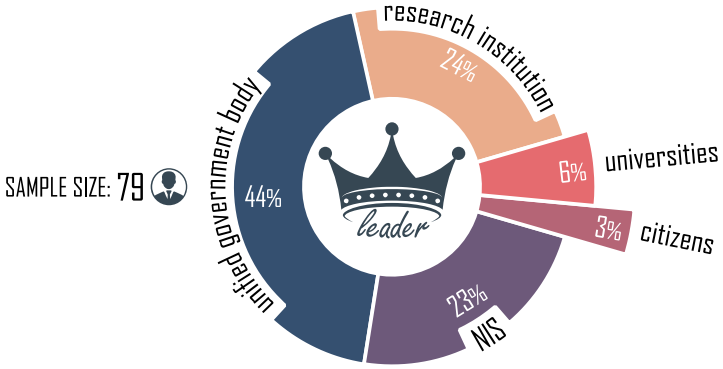


Figure 7: Institutions to take leading role in expanding the open data community

Because of this infrastructure deficiency, many interviewees (44%, Figure 7) expressed their wish that a unified government body should be mandated to lead in expanding the open data ecosystem so that a single point-of-access to government data could be created to streamline data discoverability.

Apart from the government's websites, this survey also looked at open data infrastructures operated by the UN's various bodies and other international institutions due to their crucial roles in data collection, production, dissemination and consumption in their capacity as Cambodia's major development partners and policy influencers. It appears the UN has made more data about Cambodia available to the public than the Cambodian government itself. For example, the UNData database has made 1,560 datasets about Cambodia available for public consumption (row 2 of Table 10 in the Appendices). The Humanitarian Data Exchange, another UN data portal, provides the public with access to 119 spatial datasets about Cambodia (row 1

of Table 10). Although the UN still does not have a single point-of-access, the way the data is organized in each of its websites is much more user-friendly. They are more systematic and structured and offer data search functionality which greatly increase data visibility. Moreover, the data is always presented in various machine-readable formats.

Besides the UN, several international development partners also provide access to their data. However, most of the data is related to the implementation of their development projects. Among these international organizations, the Mekong River Commission holds the largest number of datasets (2,903 datasets) about Cambodia (row 12 of Table 11 in the Appendices). However, these datasets are not always open, and a large number require the purchase of access privilege. These international bodies have built a good digital data infrastructure such as data portals or queryable web databases that provide workable metadata, data collection methods and the ability to retrieve the data in bulk.

Cambodia's civil society provides almost no open data infrastructure. The survey team reviewed more than 30 websites of CSOs that were identified as important data producers due almost exclusively to their project portfolios and reputations. We found that, except for a few well-known NGOs such as ODC which holds 221 datasets (as of Feb. 14, 2020), LICADHO (two datasets on land concessions and court watch as of Feb. 10, 2020), WCS Cambodia (dataset on biodiversity of Keo Seima Wildlife Sanctuary), most of them, from an open data perspective, did not provide machine readable datasets. The large majority of the CSOs' databases do not contain data, and the minority that do provide access to their data do not adhere to open data standards. Data in these websites (Table 9 in the Appendices) is aggregated in summary reports. A similar situation exists regarding the Cambodian universities and research institutions. None of these establishments have built any open data infrastructure. For example, the Cambodia Development Resource Institute (CDRI) has made no data open despite boosting a large portfolio of research projects. Although its many research reports are available free of charge on its website, no associated open data is provided.

5.5 Open Data Literacy and Attitude

Although this survey did not conduct any comprehensive assessment of data literacy and attitudes towards open data, it was able to collect some anecdotal evidence that indicated low data literacy levels and worrisome attitudes towards open data stewardship and usage. Although 88% of the study's participants acknowledged that open data is incredibly important for their work, career, and the development of Cambodia in general, only three percent correctly defined what constitutes open data. Although most of them correctly identified data availability (53%) or accessibility (74%) as what make data open, these definitions are only partially correct. In parallel to what we found from the survey questionnaire; all journalists interviewed intimated that they did not know what constitutes open data. They also claimed that their colleagues in the journalism sector were not trained to work with complicated data and lack functional data literacy.

What is more alarming is that university researchers who depend regularly on data from external sources were not themselves aware of the licensing schemes being placed on the data they use. When asked whether they know about the Creative Common licensing schemes used by ODC to govern their released data, they were completely surprised to learn that open data is also governed by licenses. This finding is also in line with the fact that a large proportion of the questionnaire respondents (45%) do not know what metadata–a set of data that describes and

summarizes basic information about the corresponding data–means (Figure 8).

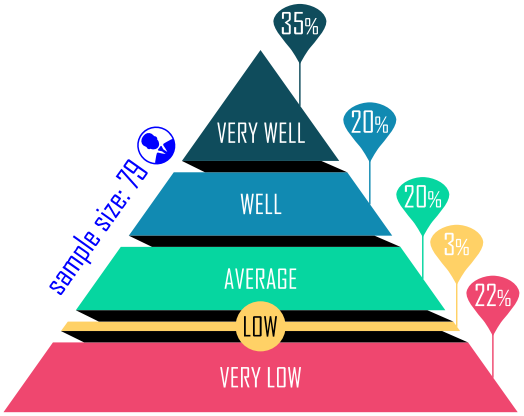


Figure 8: Level of understanding of metadata

5.6 Open Data and Cambodia's Regional Position

Realizing the importance of open data, many countries have been developing national data portals in many different forms to encourage citizen engagement in the government's development efforts. Developed countries such as the US, the UK, Canada, and Australia have built sophisticated open data portals. The US, for example, has a very advanced and highly distributed Data Portal that connects users to 197 different data custodians and holds an extensive collection of 254,657 datasets (as of Feb. 25, 2020). Acting as a 'front end', the portal does not host data directly but rather harvests data held independently by each participating organization who runs their own compliant data repository. In a similar fashion, the Australian national data portal is interfaced to the public through its data.gov.au site. For the UK, the public can access its open data portal at data.gov.uk site. The gateway to Canada's portal is its "Open Map" site. These are some examples of functioning portals in highly developed countries. In developing countries, the progress of open data development is much slower. Although the Indian government, under a joint initiative with the US government, runs a national Open Data Portal, the launch of the National GIS initiative in 2011 signals at the fragmented functioning of its open data infrastructures.

Table 3: Data portals of selected ASEAN countries

Country	Portal	Datasets**
Brunei	data.gov.bn	598
Indonesia	Satu Data Indonesia	43,095
	*Ina-Geoportal	5,671
Malaysia	Open Data Portal	13,612
	*Malaysia Geoportal	5,908
Philippines	Open Data Philippines	278
	*Philippine Geoportal	949
Singapore	data.gov.sg	1,801
Thailand	data.gov.th	1,330

*Geoportal **As of Feb. 25, 2020

For the ASEAN bloc, Cambodia, Laos, Vietnam, and Myanmar do not have any functioning government data portals. Other members have built open data platforms (Table 3). It is of note that some of these countries also operate separate geoportals. These portals provide

data discovery and access to the public. However, a fully functional open data portal requires that accessible data be processed or analyzed in an integrated way. The platforms in Table 3 did not necessarily accommodate this due in large part to incomplete metadata. Information such as scale and resolution, capturing technique, and limitations of the data is not always available. Some of these portals also only provided metadata of the actual datasets, which are in turn kept offline.

This discussion indicates that Cambodia is far behind its neighbors in terms of open data development. It is being surrounded by larger economies which are investing heavily on opening underused government resources to the public to spur innovations and facilitate oversight.

5.7 Challenges Faced by the Open Data Community

The deficiency of enabling open data infrastructures and low data literacy combine to create an incredibly challenging ecosystem for the open data community. Information gaps, unreliable data due to poor data documentation and metadata, and low data quality are the top three perceived challenges faced by the data users (Figure 9**Error! Reference source not found.**). While many of these challenges are interrelated, they are the consequences of the lack of infrastructure and low data literacy.

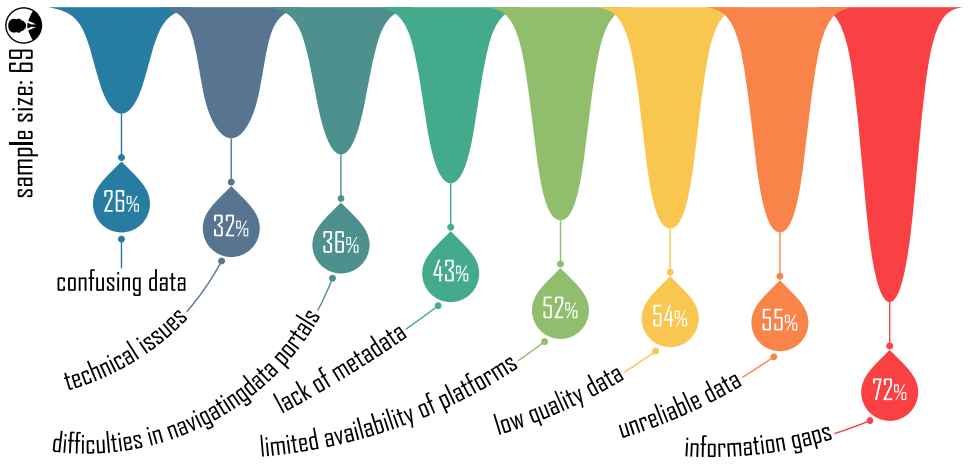


Figure 9: Challenges in using open data

The lack of data literacy is probably the main contributing factor towards low data quality. The survey reviewed several datasets hosted at ODC and the Humanitarian Data Exchange and found that documentation related to data creation and update process was somewhat incomplete. Most datasets did not meet metadata standards. Changes to the datasets through their life cycles also were not sufficiently recorded, resulting in possible confusion and data inconsistency. For example, many datasets hosted at the Humanitarian Data Exchange listed "*Direct Observational Data/Anecdotal Data*" as the data creation method. This phrase does not provide reliable information on how the data was captured, what equipment was used, what accuracy level or scale was, etc. Changes to the data because of subsequent updates were not documented at all. In the case of ODC, the data compilation process was mostly documented. However, because ODC generally compiles data from external sources and rarely conducts data collection by itself, the quality of its data depends greatly on the original sources. The lack of

recognizable and reliable data portals also encourages informal data exchanges that are prone to data integrity problems.

The challenges in using the data create many direct difficulties that hinder the growth of the open data community. As evidenced by the facts presented in Figure 10, the relationship between challenges in data usage and in community growth is clearly established.

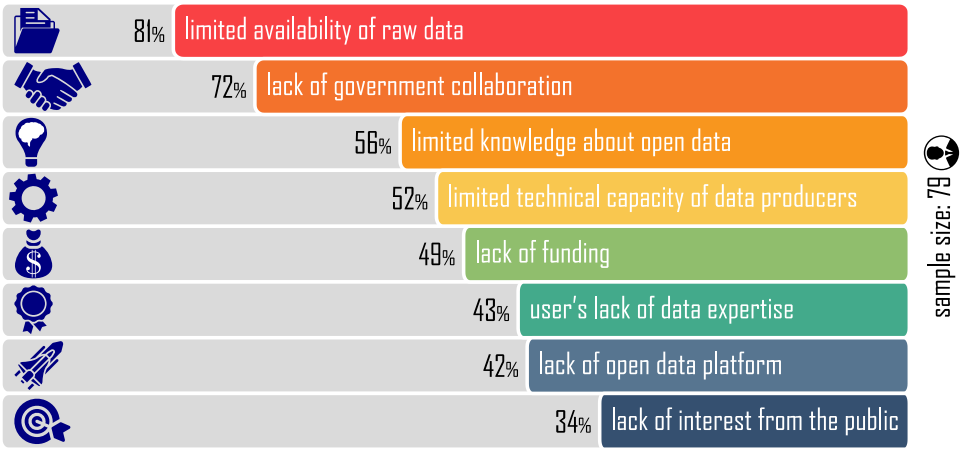


Figure 10: Challenges hindering the growth of the open data community

6 Implications for Cambodia's Development

The findings presented in section 5 have profound implications on Cambodia's development. This section presents the survey team's assessments of how a weak open data ecosystem impacts the development of Cambodia. These assessments are the synthesis of this survey's findings and the knowledge and experience of the authors. They are, therefore, viewpoints seen from the lens of an open data professional and a socio-economic development worker.

6.1 Data Gaps and Development Challenges

From Figure 9 and Figure **10**, information and data gaps are the two main factors affecting the open data ecosystem. These two challenges complement each other in a cyclic way. Poor-quality data leads to incomplete, incorrect, or misleading information which in turn results in bad decisions and actions, which eventually lead to the creation of new poor-quality data and thereby complete a cyclic revolution. Decision and policy makers who depend on accurate and reliable data for timely decisions will suffer from the unavailability of quality decision inputs. And in the case of Cambodia, they are related in a large part to the education, environment, health, and infrastructure sectors, which are the main driving forces for sustainable development (refer to Figure 11). These four sectors comprise the backbone of any developing economy. The lack of quality environmental data can lead to poor decisions on natural resources extraction initiatives that are socially and environmentally unsustainable. Human development will inevitably suffer from lack of information on the education, health, and infrastructure sectors. But the biggest data-induced problem that crosscuts every development aspect is that the limited availability of open data undercuts the public's engagement with and oversight of the government's development initiatives. The public, as rights holders, need accurate, timely and reliable data to become well-informed and empowered. An intelligent and empowered public has a much better chance of holding the government, as the duty-bearer, accountable and

thereby creates an inclusive development environment.

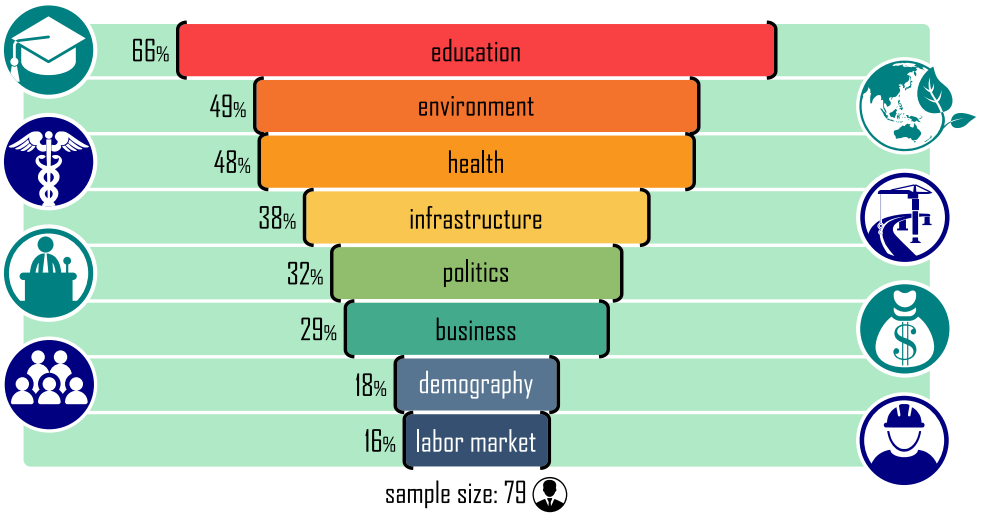


Figure 11: Perceived information gaps in different sectors

6.2 Open Data and Regional Competition

The potential impacts of open data on the development of Cambodia cannot be viewed from a single and isolated perspective. Real development challenges come from both internal and external factors. The discussion presented in section 5.6 clearly highlights the development pressure induced by regional forces on Cambodia's competitiveness, creativity and innovation. With bigger neighbors opening up their data repositories to their public and investing heavily on enabling data infrastructures, one would wonder how the Cambodian people will deal with a fast-changing, data driven digital economy and remain competitive. The rapid spread of digital technologies is highly disruptive to the traditional workplace and has altered the dynamic of the labor market. A cheap and unskilled labor force that has carried Cambodia's sustained economic growth for the past two decades[6] has come under mounting pressure caused by competition from automation and the ICT-based job market. However, rising wages and decreasing trade preferences have a real potential to derail Cambodia's economic development. The exploding adoption of mobile technologies that results in better access to information for the public has consequently empowered Cambodia's young and vibrant population that is increasingly more imposing socially, economically, and politically. The dilemma surrounding these data driven changes for the Cambodian government centers around how well-prepared it is to invest in data production and delivery systems that ensure a reliable flow of data to the hands of its people who will then turn it into innovative and creative products and services. Failure to seriously invest in data infrastructures and literacy will greatly reduce Cambodia's competitiveness in the face of increasing competition from its neighbors.

6.3 Opportunity for Open Data Growth

Despite these many issues, some opportunities for the growth of open data exist. An example is the World Bank funded Higher Education Institutions Capacity Improvement Project

[6] World Bank, *Benefiting from the Digital Economy: Cambodia Policy Note*, 2018. Accessed Jan. 21, 2020

(HEICP),[7] under which many scientific research projects will be supported. These projects will inevitably require a lot of open data and eventually will drive huge data growth. Another example is that the Cambodian government has begun to appreciate the importance of open data portals, as demonstrated by the launch of the Ministry of Environment's data portal. All the researchers interviewed have shown a great interest in the concept of Volunteered Geographic Information (VGI) which depends on crowd-sourced open data to produce products and services. Two notable examples of this concept are the OpenStreetMap and FlightRadar24 projects. But most importantly, the immature open data ecosystem in Cambodia points to a huge potential for growth as a matter of necessity to fill the information gaps discussed above.

But the most important opportunity among all is perhaps related to the potential held within NCDD's databases, whose coverage reaches the most grass-roots levels. Increasing public awareness of these databases and how they can be used to increase public accountability will open the door to a much better citizen engagement.

7 Ways Forward

The survey uncovered several aspects of the open data ecosystem in Cambodia which can serve as good starting points for future interventions. The following discussion suggests some key recommendations drawn exclusively from this study that can be used to build a more vibrant open data community in Cambodia.

1. Institutionalize a unified government body to facilitate open data development. This survey has shown that the institution perceived as most capable to take the leading role in expanding the open data community is a unified government body (Figure 7). The biggest challenge in open data development is to bring in all different data producers under one single framework. Having a unified body that can create strong institutional arrangements such partnership policy, coordination and financing and technical uniformity such as interoperability standards, data specifications, and quality control mechanisms will ensure that all data producers are on the same communication channel, talk a single technical language, and share a common understanding of joint benefit sharing and prosperity.
2. Build a strong network of open data infrastructures: Up to this point, there is no doubt that a vibrant open data community will require strong enabling infrastructure. Although they may consist of access networks, cloud servers, web services, and data centers and other important hardware, the most visible of these enabling infrastructures is the data portal, or the front-end to the open data universe that connect data users to data producers. Building robust open data portals is seen thereby as imperative for the growth of the open data community.
3. Increase data literacy among both data users and producers: If the two recommendations above are vital for the prosperity of the open data community, this last recommendation binds them together. Low data literacy is what drives the public's apathy towards open data development. When journalists cannot make use of complicated data, data users do not understand licensing, and data producers fail to value data documentation and metadata, a good data portal and unified regulatory body will only get us halfway. The human side of the open data community must be taken seriously into account. A highly data literate public is data creative and able to transform data into products, services, and decisions.

[7] Cambodia and World Bank Join Forces to Improve Higher Education and Connectivity, press release, Jul. 12, 2018. URL:https://www.worldbank.org/en/news/press-release/2018/07/12/cambodia-and-world-bank-join-forces-to-improve-higher-education-and-connectivity. Accessed January 21, 2020

Appendices

Table 4: Government ministries with published data

Institution	Data	Remark
1. Water Resources and Meteorology	• Automatic weather records • Meteorological data • Flood information and hydrographs • Flood and drought maps	• Data is provided as web maps, map images
2. Post and Telecommunication	• Basic data on telcom and postal operators and infrastructures	• Data is provided in summary reports as PDF documents
3. Economy and Finance	• Budget Brief • Government Finance Statistics • Economic and Financial Statistics Bulletin • Consumable Market Price • Inflation Report • Cambodia Public Debt Statistical Bulletin	• Data is given in aggregated forms in PDF formats
4. Education Youth and Sports	• Education Statistics and Indicators	• Data is provided as PDF documents
5. Land Management, Urban Planning and Construction	• List of registered local architectural and construction companies • List of registered foreign architectural and construction companies • List of registered architects and civil engineers	• Data is provided as PDF documents
6. Planning	• National statistics	• Refer to the National Institute of Statistics (NIS)
7. Culture and Fine Art	• Refer to Apsara Authority	
8. Environment	• Refer to NCSD	
9. Agriculture Forestry and Fisheries	• Rice Production Statistics • Price of Agricultural Products • Business Network Database	• Data is given as web and PDF documents
10. Rural Development	• Rural Road Information Management System	• Restricted internal system
11. Commerce	• Special Economic Zone • Commodity Price Bulletin • Trade Statistics • Business Registration Entity Search	• Data is given in PDF documents and Microsoft Excel files
12. Civil Service	• Staff Information Management System • Document Management System	• Data usage is restricted to registered users
13. Industry and Handicraft	• Water Supply Monitoring System	• Restricted internal monitoring system

Table 5: Notable sub-ministerial government agencies without published data

Institution	Institution
1. Foreign Affairs and International Cooperation	2. Religions and Cults
3. Justice	4. Tourism
5. Social Affairs Veterans and Youth Rehabilitation	6. National Defense
7. Mines and Energy	8. Health
9. National Assembly Senate Relations and Inspection	10. Public Works and Transport
11. Information	12. Women's Affairs
13. Interior	14. Labor and Vocational Training

Table 6: Sub-ministerial government agencies with published data

Institution	Data	Remark
1. National Audit Authority	• Annual audit reports	• Aggregated data inside the report
2. National Election Committee	• Voter list • Election results	• Data is provided as searchable web list and PDF documents
3. National Bank of Cambodia	• Monetary and Financial Statistics • Economic and Monetary Statistics • Balance of Payment • Banking data • Microfinance Institution data	• Data is given in Microsoft Excel or PDF formats
4. National Institute of Statistics	• Consumer Price Index (2019) • National Indicator Reporting Platform • National Account • National Data Archive (NADA) • National Data Archive (30 licensed datasets, 3 public datasets)	• Data is given in HTML format
5. National Committee for Sub-national Democratic Development	• Project Implementation Database • Gazetteer Database Online • Commune Database Online • Commune Council Database • Bidding Announcement Database • ISAF M&E Database System • DMK M&E Internal System • Sub-national Project Database	• Commune Database Online was updated to 2010 • Commune Council Database was updated to 2008 • Sub-national Project Database does not seem to work
6. National Committee for Disaster Management	• Disaster Loss Database	• Queryable web forms (charts, statistics, maps) with ability to export data as CSV format
7. The National Council for Sustainable Development	• Species Database	• Data is given as web tables without export option

Table 7: Sub-ministerial government agencies with published data (cont.)

Institution	Data	Remark
8. National Road Safety Committee	• Road Accident Database	• HTML table (no data export option) and non-spatial data
9. General Department of Customs and Excise	• Automated System for Customs Data	• Restricted internal system
10. Cambodia Securities Exchange	• Market Data (stock, bond, index)	• Proprietary data
11. Secretariat of Civil Aviation	• Statistics • Vital data	• Data is inaccessible possibly due to broken URL links
12. Securities and Exchange Commission of Cambodia	• Market participants	• Web documents
13. APSARA Authority	• Tourist statistics	• Data is given in scanned PDF files. Some links are not working
14. Sihanoukville Autonomous Port	• Corporate disclosure	• Links to report are not working
15. Telecommunication Regulator of Cambodia	• Telecommunication statistics	• Data is given as summarized web charts

Table 8: Notable sub-ministerial government agencies without published data

Institution	Institution
1. Cambodia National Mekong Committee	2. Council for the Development of Cambodia
3. Anti-Corruption Unit	4. National AIDS Authority
5. National Accounting Council	6. General Department of Taxation
7. Constitutional Council of Cambodia	8. General Department of National Treasury
9. The Council of Ministers	10. Electricity Authority of Cambodia

Table 9: Notable local CSOs with published data

Institution	Data	Remark
1. LICADHO	• Land concessions • Court watch	• Data is given as web documents or downloadable files
2. Transparency International	• Aggregated data	• Data summary in research reports
3. Cambodia Development Resource Institute	• Aggregated data	• Data summary in research reports
4. NGO Forum on Cambodia	• Aggregated data	• Data summary in research reports
5. Save Cambodia's Wildlife	• Aggregated data	• Data summary in research reports
6. WCS Cambodia	• The biodiversity of Keo Seima Wildlife Sanctuary	• Downloadable dataset
7. Open Development Cambodia	• 221 datasets	• Downloadable datasets in CSV format

Table 10: Notable UN bodies with published data

Institution	Data	Remark
1. The Humanitarian Data Exchange	• 119 datasets	• Data is provided in many downloadable formats
	•	• Data is accompanied by usage license
2. UNData	• 1560 datasets	• Statistical data in several downloadable formats
3. SDG Indicators Global Database	• 5271 observation	• Data is downloadable in several formats
4. FAOStat	• Data covering 10 topics	• Data in downloadable CSV format
5. ILOStat	• Data on various topic	• Data in several downloadable formats
6. WHO Health Information and Intelligence Platform	• Data on various topic	• Data in downloadable Excel format
7. Global Health Observatory	• Data on various topic	• Data is given as image files
8. International Monetary Fund	• Various databases	• Web and downloadable data
9. UNIDO Statistics Data Portal	• Various databases	• Require registration for access
10. International Telecommunication Union	• Various indices	• Downloadable PDF
11. World Bank Data Catalog	• 92 Datasets	• Data pulled from different global databases
	•	• Survey datasets are also provided
12. UNDP Human Development Database	• Data on 14 dimensions	• Data is given as web charts and downloadable CSV files
13. UNESCAP	• Country development indicators	• Web resources and downloadable files
14. UNICEF	• Various databases	• Data can be downloaded as CSV files
15. UNHCR	• Various databases	• Data can be downloaded as CSV files
16. UNESCO	• Various databases	• Data can be downloaded as CSV files
17. WFP GeoNode	• 10 layers of spatial data	• Data is given as Web Map Services

Table 11: Notable international bodies with published data

Institution	Data	Remark
1. Demographic and Health Survey	• Five datasets	• Access requires registration
2. Enterprise Surveys Dataset	• Four datasets	• Access requires registration
3. Asian Development Bank	• 54 datasets	• Downloadable in different formats with usage license
4. Oxfam	• Program data	
5. Japan International Cooperation Agency	• Aggregated data	• Data summary in reports
6. Korea International Cooperation Agency	• Statistical data	• Data is summarized in annual statistical report
7. USAID	• 50 datasets	• Downloadable data and registered-user-only access
8. International Aid Transparency Initiative	• 29 datasets	• Downloadable data in different formats
9. WWF	• Conservation Science Data and Tools	• Various downloadable datasets
10. IUCN	• IUCN red lists	
	• World Database on Key Biodiversity Areas	
	• World Database on Protected Areas	• Open or restricted database
11. The Asia Foundation	• The Asia Foundation Data Portal	• Data on survey results
		• Data access through request
12. Mekong River Commission	• Data portal containing 2903 datasets	• Datasets are available without restriction or require purchase