

## **An Open Science ‘State of the Art’ for Hong Kong: Making Open Research Data Available to Support Hong Kong Innovation Policy**

Naubahar Sharif<sup>1</sup>  
Waltraut Ritter<sup>2</sup>  
Robert L Davidson<sup>3</sup>  
Scott C Edmunds<sup>4</sup>

*Open Science is an umbrella term that involves various movements aiming to remove the barriers to sharing any kind of output, resources, methods or tools at any stage of the research process. While the study of open science is relatively advanced in Western countries, we know of no scholarship that attempts to understand open science in Hong Kong. This paper provides a broad-based background on the major research data management organisations, policies and institutions with the intention of laying a foundation for more rigorous future research that quantifies the benefits of open access and open data policies. We explore the status and prospects for open science (open access and open data) in the context of Hong Kong and how open science can contribute to innovation in Hong Kong. Surveying Hong Kong’s policies and players, we identify both lost research potential and provide positive examples of Hong Kong’s contribution to scientific research. Finally, we offer suggestions regarding what changes can be made to address the gaps we identify.*

*Keywords: Open Data, Open Access, Open Science, Hong Kong, Innovation Policy, Research Data Management*

### **Introduction**

The OECD defines open science as policies that aim ‘to make the primary outputs of publicly

---

1 Division of Social Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR.

2 Knowledge Dialogues, Hong Kong SAR; Open Data Hong Kong, Hong Kong SAR.

3 Open Data Hong Kong, Hong Kong SAR; TwoCloaks Ltd, Glasgow, United Kingdom.

4 Corresponding author. Open Data Hong Kong, Hong Kong SAR; GigaScience, BGI HK Limited, 16 Dai Fu Street, Tai Po Industrial Estate, Hong Kong SAR. scott@gigasciencejournal.com.

funded research results—publications and the research data—publicly accessible in digital format with no or minimal restriction’ (OECD, 2015). Open science extends the principles of openness to the whole research cycle, fostering sharing and collaboration as early as possible and is a systemic change to the way science is done. Open science is frequently used as an umbrella term that encompasses various movements aimed at removing the barriers to sharing any kind of outputs, resources, methods or tools at any stage of the research process. As such, open access to publications, open research data, open source software, open collaboration, open peer review, open notebooks, open educational resources, open monographs, citizen science and research crowdfunding all fall into the category of open science. The focus is usually placed on open research data and open access to scientific publications.

Across the globe, governments are releasing increasing amounts of what has become known as ‘open data’. Because much academic research is government funded, policies and practices that encourage open data often apply to publicly funded research outputs, though there are vast differences across localities. For instance, governments in North America and Europe view open data as a common good and make efforts to improve access to open data and to more effectively use the results of government-funded research.

While the study of research data management and policies is relatively advanced in Western countries, we know of no scholarship that attempts to understand Hong Kong’s research data management policies. This paper fills this lacuna by providing a broad-based background on the major research data management organisations, policies and institutions with the intention of laying a foundation for more rigorous future research that quantifies the benefits of open access and open data policies. In this paper, we explore the status and prospects for open science (open access and open data) in the context of Hong Kong and how open science can promote innovation in Hong Kong. We detail how the rest of the global research community carries out open science, contrasting the policies and strategies of the leading countries in this area with those in Asia and Hong Kong. Surveying Hong Kong’s policies and players, we both identify lost research potential and provide positive examples of Hong Kong’s contribution to scientific research. We outline the public policies, using freedom of information (FOI) requests where these are unclear, and survey and interview various stakeholders to fill in many of the remaining gaps. Finally, we offer suggestions of what changes can be made to address the identified gaps.

The remainder of this paper is structured as follows. We first offer a global picture of open science before moving on to examining open research data policies in Asia. This is followed by a more focused look at Hong Kong when we introduce the situation and players there. The methods we use for our research are then introduced, and our findings are discussed with stakeholder perspectives. Based on these findings, we draw conclusions and offer recommendations.

## **Open Science: The Global Picture**

In its influential 2012 ‘Science as an Open Enterprise’ report, the Royal Society highlighted the need to tackle the deluge of data created by digital technologies. Preserving the principle of openness and exploiting data in ways that maximise its re-use has enormous potential for science and its application in public policy and business. Greater accessibility to scientific data can be incentivised through greater recognition of the value of data gathering, analysis and communication. To enable such an outcome, the Royal Society stated that ‘intelligent openness requires that data must be readily discoverable and accessible; made intelligible to those who may wish to scrutinise it; and provided with sufficient metadata and background material to make it assessable and usable’. Open access to research data carries cost implications requiring trade-offs that reflect the value of money and use (Royal Society, 2012).

Funding agencies such as the NIH and the NSF in the US and the European Union Framework Programmes, following the lead of independent funders such as the Wellcome Trust and the Gates Foundation, have been leading the way in mandating open access to published research outputs and have been developing parallel policies specifically for data access.

In the UK, the Expert Advisory Group on Data Access, which represents and advises four of the UK’s biggest funders, has published detailed reports and statements encouraging scientists to voluntarily make their data more readily available to others and has made it clear that funders need to provide support and resources to enable this (Bobrow, 2015). In the US, to guide future efforts to improve access to the results of federally-funded research, the National Science and Technology Council’s Committee on Science established the Interagency Working Group on Open Science. As a result, 22 federal departments and agencies, accounting for more than 99 percent of US federal R&D expenditure, now have public access plans in place for federally funded publications and data (The White House, 2017).

These moves are being mirrored in many other countries, particularly in Australasia, with the Australian government’s Chief Scientist publishing a report in 2014 titled ‘STEM: Australia’s Future’. The report states that ‘publicly-funded research outcomes must be recognised as a public good, and also be widely diffused or used commercially in ways that benefit Australians. Researchers should be supported to share data, communicate findings and contribute to the global stock of knowledge’ (Office of the Chief Scientist, 2014).

## **Open Research Data Policies in Asia**

Asia accounts for close to half of global expenditure on research and development (42%) (UNESCO, 2015), but as a region lags far behind its Western counterparts in sharing the outputs of this research (Shearer, Repanas, Yamaji & COAR Community of Practice, 2017). While governments and their research funders in Europe (ERAC Secretariat, 2016) and North America

(Holdren, 2013) have made statements that the work of universities and the research and data outputs that they fund are for the public good, that is not part of the discourse in Asia.

There is evidence that a lack of supporting data reduces re-use and subsequent citation of research outputs (Piwowar & Vision, 2013) and governments and funders cannot maximise their research budgets without following open policies and practices. While Mainland China has recently begun moving towards adopting open access and research data policies, Hong Kong has yet to follow suit.

In 2014 the National Natural Science Foundation of China (NSFC) and the Chinese Academy of Sciences (CAS) announced that researchers they support should deposit their papers into online repositories and make them publicly accessible within 12 months of publication (Van Noorden, 2014). Mainland China has also started to make progress in the area of data policies, with the National Science and Technology Infrastructure Program requiring national research projects they fund in areas including meteorology (CMA, 2017), earth science (Geodata.cn, 2017) and agriculture and forestry (China Forest Science Data Centre, 2003) to submit and share data. The Chinese Ministry of Science and Technology (MOST) recently announced new policies and the formation of a new science data centre that will be ‘promoting open access to, and sharing of, science data’ (MOST, 2018).

In India, the Departments of Biotechnology and Science & Technology have jointly adopted an open access policy for publicly funded research (DBT India, 2014) and the Wellcome Trust/DBT India Alliance (the ‘India Alliance’) supports ‘unrestricted access to the published output of research’ (Wellcome Trust DBT India Alliance, 2016).

The Singaporean Agency for Science, Technology and Research (A\*STAR) developed its open access mandate in 2014 (ROARmap, 2014a), and while Singapore currently does not have policies regarding research data, they are looking closely at the UK model and Concordat.

Taiwan has only one university with an open access policy, National Chung Hsing University (ROARmap, 2014b) and it serves as an example in Asia of mandated open data policies for research data. Taiwan’s Ministry of Science and Technology has encouraged and funded the building of data resources in various life sciences fields (Taiwan MOST, 2018). The ministry mandated that the Institute of Oceanography release all data produced since 2011 in its Ocean Data Bank repository (ODB) (Taiwan MOST, 2014). Taiwan’s Council of Agriculture has similar policies in place for terrestrial ecology studies funded by the Forestry Bureau (Taiwan Forestry Bureau, 2016).

## **The Situation and Players in Hong Kong**

The Hong Kong Research Grants Council (RGC) is committed to sharing the research findings from studies it funds with the public for projects approved from 2010/11 onwards. The principal investigators of completed projects are obliged to release completion reports highlighting research outputs such as publications, conference presentations and students. While there is also an ‘Other impact’ section that includes patents, prizes, collaborations with other research institutions and technology transfer, there is no mention of data or software that is equivalent to how the NSF encourages data citations in its fund recipients’ bio-sketches (RGC, 2011). Apart from a small number of projects known as joint research schemes and a short layman’s summary, none of this information is currently available to the public.

Another government agency in Hong Kong that can advocate for the sharing of research data is the Innovation and Technology Commission (ITC). The ITC administers the Innovation Technology Fund (ITF), which provides financial support to projects that contribute to innovation and technology upgrading in Hong Kong, including both R&D and non-R&D projects. The ITF guidelines require fund recipients to make R&D outcomes available to industry and the public, but only ‘aggregated summary information’ of funded projects, including project abstracts, deliverables, implementation organisations and contact persons, are required to be published on the website dedicated to the ITF (Innovation and Technology Commission, 2017).

For the majority of the projects funded by the ITF and conducted by universities or R&D centres set up by the government or other designated research institutions (except those which are proprietary in nature), fund recipients are required to make their R&D outcomes available to the public to facilitate commercialisation and the transfer of information to industry, such as through non-exclusive licensing arrangements.

While the RGC and ITF are the highest profile research funders in Hong Kong, other government funders of research include the Food and Health Bureau (FHB), which administers many grants under the Health and Medical Research Fund, the Central Policy Unit (CPU), which funds public policy research, the Development Bureau (DB) Built Heritage Conservation Fund and the Environment and Conservation Fund. There is some overlap between these funders and the UGC/RGC, but they perform as little or less scrutiny and guidance of the research outputs resulting from the projects they fund. According to information gained through FOI requests, the Health and Medical Research Fund and the Environment and Conservation Fund require final evaluation reports but have no policies on what happens to research outputs such as publications and data (Environment Bureau, 2017; Food and Health Bureau, 2017).

The latest player in this hierarchy is the Innovation and Technology Bureau (ITB), founded in November 2015 with the responsibility for policy matters on innovation and technology. The ITB has an Innovation and Technology Branch and oversees the operation of the ITC and the

## Office of the Government Chief Information Officer (OGCIO)

Figure 1 depicts the roles played by these and other organisations and stakeholders in Hong Kong's science and technology infrastructure, providing an overview of the complex interlocking institutional system through which information and technology moves from the laboratory to the world of practical applications.

**Table 1 Hierarchical Representation of S&T Actors in Hong Kong**

<b>Political Forum</b>	Legislative Council (LegCo)				
<b>Policy- Makers</b>	Government		Advisory Committee on Innovation and Technology		
	Innovation and Technology Bureau (ITB)		Innovation and Technology Commission (ITC)		
<b>Financing</b>	Government		Education Bureau (EB)	Private Sector	
	ITC → Innovation and Technology Fund & IT Venture Funds (ITF & IT VF)	Other Gov Funders	Research Grants Council	University Grants Committee	
<b>Operators</b>	Universities	Public Technology Support Organizations		Private Sector	
	R&D Centres	Applied Science and Technology Research Institute			
<b>Facilitators</b>	HKPC	HKTDC	HKSTPC	Cyberport	HKIB
<b>Commercial-ization Agents</b>	Business Enterprises			New High Tech Ventures	Multinational Corporations

*Abbreviations:* HKPC—Hong Kong Productivity Council  
HKTDC—Hong Kong Trade Development Council  
HKSTPC—Hong Kong Science and Technology Parks Corporation  
HKIB—Hong Kong Institute of Biotechnology

IT VF—Innovation & Technology Venture fund

Other Gov— Other government funders including FHD, CPU and DB.

## Methodology

For this study, we contacted the primary funding bodies in Hong Kong, the UGC and the RGC, to request any data on which funded projects had been published in open access journals as a percentage of total publications. The RGC operates under the aegis of the UGC and functions as an advisory body on research matters within the organisational structure of the committee. Following its response, we contacted the research offices of the eight publicly funding Hong Kong universities to request the corresponding statistics from their own organisations. Here we present selected responses from each institution to indicate broadly the status of and attitudes towards monitoring and encouraging open access publishing and open data among publicly funded researchers.

In addition to using UGC- and RGC-funded institutions' data, we used the following four data sources for this paper. First, we examined the scholarly literature pertaining to the development of Hong Kong's innovation system and economy and the international literature on managing and sharing research data. Second, we performed in-depth analysis of official Hong Kong government documents and supranational documents from the OECD, the EU and the Royal Society. Third, we gathered information pertaining to the functions of government departments from their official websites. Where information was not publicly accessible, we submitted requests for information based on the Hong Kong Code on Access to Information (HKSAR Government, 2016). These are archived and accessible via the accessinfo platform. Finally, to fill any remaining gaps, we interviewed relevant stakeholders both within the Hong Kong government and non-government individuals who have had (and continue to have) an important role in shaping policymaking in this domain in Hong Kong.

Surveys and interviews to explore current practice in research data management were conducted with major stakeholders in Hong Kong, including all UGC-funded universities, individual professors, the ITC, ASTRI, R&D Centres, Citizen Science groups and the UGC itself. We also emailed all librarians and research officers from all 8 Hong Kong universities (with responses in both cases from 5/8 of them).

While the list of people we approached is not exhaustive, it does capture the core group of individuals associated with research data management policies in Hong Kong. To understand the origins, use and future of research data management policies, we had more detailed conversations with the individuals listed in Table 1 between January 2014 and March 2016. These interviews were semi-structured, using a guide we constructed to direct conversation. Where possible, these guides were provided in advance to interviewees but did not necessarily articulate the exact wording of the questions that we would ask or the order in which we would

raise particular issues, rather providing an outline of themes that we wanted to cover during the interview. This technique helped ensure that we covered all issues of particular interest while still allowing flexibility to tailor the interview to the responses of the interviewees. We made handwritten notes during the interviews, but mostly focused on the oral discussion taking place. Each of the interviews lasted between 45 minutes and two and a half hours. The Additional Supporting Data archived in OSF presents summary information on the interviewees, including their institutional affiliations at the time of interview and their role in shaping Hong Kong's research data management policies. The data from individuals we contacted who did not agree to an interview are not recorded.

## **Findings and Analysis**

Hong Kong's UGC has an annual budget for grants of 17.5 Billion HKD (4% of government spending), funding most research at the eight Hong Kong universities. In 2011-2012, the institutions UGC funded published 16,594 papers (including conference papers and non-refereed work). With estimates that open access can increase citation impact by 50% (Piwowar & Vision, 2013) and that data deposition leads to a 9% citation boost (Harnad, 2005), open access and data policies appear to be good investments.

Our requests to funding and research bodies for statistics on the open access status of outputs received limited responses. Our initial contact with the UGC served only to point us towards the funded institutions themselves. The response, as evidenced in the following quote, demonstrated the primary funding body's hands-off approach, with emphasis on institutional freedom to pursue their own policy in this regard: 'The bulk of the University Grants Committee's recurrent grants are disbursed to institutions in the form of a block grant to provide institutions with flexibility in internal deployment'.

Of the eight institutions contacted, we received four responses. Each respondent stated that his/her institution did not record the open access status of its outputs. Our request specifically asked for details regarding UGC-funded projects and two of the respondents admitted that they could not provide statistics on publications sorted by funding body. The City University of Hong Kong (CUHK) specified that it would not collect this level of information because it is not required for the Common Data Collection Format (CDCF), the UGC's standard form for recording institutional performance (University Grants Council, 2015).

Initial investigations into the contents of the CDCF yielded only basic information about when it was first developed and indicated that it had undergone periodic changes, but there was no publicly available indication of current or past specifications. Our request for these documents made under the Code on Access to Information was refused, as was the request for internal



review, with the response that ‘there is a real concern that any disclosure of such advice may prejudice the frankness and quality of the advice from institutions in future review exercises of the CDCF which will have a material effect on the usefulness of the reviews’ (University Grants Council, 2015).

The available data make it apparent that the CDCF does not operate with either the concept of open access or the concept of open data, software or other modern output types beyond the catch-all ‘other’ output category. It seems that this would be the logical place to capture this information if it were recorded.

*Role of Policymakers (RGC) and Implementers (Universities):* The challenge of defining a research data management policy that enhances integrity, protects privacy and benefits the public across universities in Hong Kong can (and should) be taken on by the RGC. Within the RGC, there is currently both inertia and opposition to the idea that publicly funded research is a public good. Because RGC-funded datasets are currently not stored and properly described, there is a risk of duplicating research unintentionally. As international funding agencies increasingly mandate open research data as best practice, RGC eventually might adopt common practices from other places rather than developing its own. HKU increasingly follows the ‘Oxford Principles’ for research data management (RDM Delivery Group, 2013).

This ambivalence was reflected by the chairman of the RGC, who stated in an interview that ‘there is no relationship between world-class research and release of data’, questioning whether anyone might be interested in the completeness of data. He argued that we cannot know how useful the data is or whether it is ‘good’ or ‘bad’ data. He was sceptical about the idea of making verifiable data available and asked: ‘Why show dirty laundry?’

The chairman also saw a conflict between competitiveness and openness, arguing that the reputation of a researcher is built on publications, not on the underlying data. Currently, Hong Kong universities offer no incentives or even encouragement for researchers to release their data.

Making data truly useful is a challenge. Despite the growing role of data librarians in research management, the chairman believes that most librarians lack the necessary domain knowledge to take on greater responsibility in this field.

Richard Armour, head of the RGC, responded to our email inquiry, saying that research data would not be further disclosed because such disclosures may prejudice the accuracy and quality of advice from institutions in future CDCF review exercises.

Librarians offered the most positive response to our email enquiries and survey, particularly involving their use of institutional repositories; those who responded were insightful and demonstrated that they understood the issues. The Hong Kong Education University stated that it

does not have official open access or open data policies, but its institutional repository is open to the public and has attracted roughly between 100K and 120K visits every year.

No longer waiting for UGC/RGC policies that do not appear to be on the horizon, HKU has taken initiative in 2015 by introducing a policy and taskforce on Research Data and Records Management (HKU Research Data Services, 2016). As a move towards this, ScholarHub was upgraded from an IR (Institutional Repository) to a CRIS (current research information system) to enable it to also host data, and datasets can now be indexed or hosted via ScholarHub. For an example, see HKU data in DataVerse (Liang & Fu, 2015). City University of Hong Kong (CityU) has closely watched and followed HKU's approach, recently launching a CityU ScholarHub CRIS that also allows their scholars to deposit and index datasets (Gibb, 2017).

The Hong Kong University of Science and Technology (HKUST) Institutional Repository was officially released in 2003. Indexed materials are open access scholarly materials authored by HKUST faculty, academic-equivalent staff, PhD/MPhil students and research assistants and cover journal and magazine articles, conference papers, book chapters, books, patents, theses and dissertations, presentations, working papers, technical reports, research reports and preprints. The documents in the Institutional Repository are heavily accessed; there have been 28,254 downloads over 1.4 million visits in the decade following October 2004. The HKUST repository has also been exploring the hosting of research data. This first research dataset was included in an early feasibility study on using DSpace as a data depository (Fang, Yao, Chan, Lau, & Lau, 2009).

While CUHK had no open access research data management mandate or policies regarding research data at the time of this study, its library has been involved in an awareness-raising strategy for researchers, library staff and others. The librarians surveyed were positive about and pro-active in describing their potential role in the data management lifecycle and have set up a new Research Support & Digital Initiatives team to tackle this very issue. At the time of the survey, they were focusing on handling open access articles while continuing to raise awareness of the issue of data availability. This team was up to spread the latest data-sharing protocols and platforms, aware that it needed to help researchers particularly with observation data, metadata and secondary data. New positions in the library have been set up in Scholarly Comms and Digital Services and Technologies to work on this with the aim of developing a one-stop Open Access CUHK Digital Repository for all digital contents at the university.

Finally, in terms of the Innovation and Technology Commission (ITC), an ITC spokesperson agreed that opening science and research data would help to improve the dissemination, utilisation and commercialisation of R&D outcomes. The ITC added, however, that because it does not own the intellectual property rights of the projects, it does not maintain a repository of

detailed scientific research data from the projects. In addition, because some ITF projects are proprietary in nature and involve sensitive commercial technological information, the ITC holds that it is not obligated to disseminate the research data to the public.

*Citizens Lead the Way:* While the formal research community may not be sharing its research data, much is being collected and released by curiosity-driven members of the public. So-called ‘citizen science’ is scientific research conducted, in whole or in part, by amateur or nonprofessional scientists. ‘Citizen science’ is a new term but not a new movement. The Hong Kong Bird Watching Society (HKBWS) has, for example, been publicly posting its community-collected annual birdwatch surveys since 1958 (HKBWS, 2014). The HKBWS now uses digital apps to collect geographic information system (GIS) data and data such as that collected for spoonbill sightings are available to browse on its online portal (HKBWS, 2011). Requesting access to this data is problematic, however, due to the government’s controlling copyrights for the data. The HKBWS responded to our inquiry by saying, ‘As some of the surveys were commissioned by government departments, such as the waterbird monitoring programme, the copyright belongs to the department though we are permitted to use it.’

As with birdwatching, most community interest in accessing scientific data has been in traditional hobbyist areas such as biodiversity, meteorology, geology and the environment. The WWF Hong Kong were early pioneers in this area, launching Coastal Watch in 2013 for citizen monitoring of Hong Kong’s ecologically valuable coastal habitats and continuing these in the “One Planet, Youth Action” marine litter surveys (WWF Hong Kong, 2014). The Earthwatch Institute has also operated a Hong Kong branch, organising citizen science volunteers to aid research into climate change, ecosystem services, ocean health and cultural heritage (Earthwatch, n.d.). Via FreshWater Watch, it uses volunteers to collect water and phytoplankton to understand the relationship between water quality and phytoplankton diversity in Hong Kong’s rivers. At the time of writing, over 600 datasets from Hong Kong had been collected and made publicly available via the Earthwatch portal (Stringer, 2014). Using citizen-collected data, research partners at the Open University of Hong Kong have published and presented work on red tides at international conferences (Stringer, 2014).

There is much interest in environmental and pollution data. Hundreds of radiation data timepoints have been collected in Hong Kong and are available from the Safecast crowdsourced radiation data web site (Safecast, 2016). Publicly available pollution data attracts considerable interest from downstream users, and civic hackers and software developers have produced many useful visualisations and tools using HK Observatory data, such as the Hong Kong Air Pollution App (Leyden, 2012). Branches of the Hong Kong government have finally recognised this community interest, with the Hong Kong Observatory engaging with schools and interested individuals through their Community Weather Observing Scheme (PolyU Department of Applied Physics, 2018) and the BioDiversity Division of the Agriculture, Fisheries and Conservation Department allowing the public to post pictures and register animal sightings on its biodiversity

database (AFCD, 2018) and Eco map portals (AFCD, 2015). The Census and Statistics Department has been one positive example in the Hong Kong government of a department that regularly and relatively openly shares its data, now making data available via APIs (Census Department, n.d.).

These examples demonstrate the risk that Hong Kong researchers and the government are being left behind by their own citizens, with curiosity-driven individuals collecting and making available without restriction many datasets. It may be that they demonstrate that for the public, public works are a public good, matching the RCUK position and contrasting with the RGC status quo.

## **Discussion, Recommendations and Conclusions**

In this paper, we surveyed the open access and open data situation in the Hong Kong research community, collecting policies, identifying lost research potential and providing examples of Hong Kong's contribution to scientific research and good examples that suggest the potential for successful implementation of such policies. We also offer arguments suggesting that a holistically considered open data policy can contribute positively to the promotion of innovation and innovation policy in Hong Kong.

The Hong Kong government has made strides towards disseminating public sector information, with the OGCI releasing data via its Data.Gov.Hk portal in March 2015 (previously piloted in 2011 as Data.One). Research data obtained through RGC-funded research in Hong Kong's public universities are funded by the government and can be treated in the same manner as data from the OGCI (releasing data via its portal)<sup>5</sup>.

While there are plentiful arguments in favour of open and immediate access to data, it behoves us to also examine the arguments against open data in Hong Kong. As we have already alluded to, the chairman of the RGC suggested that 'giving away' research outputs might reduce Hong Kong's competitiveness. Also, because much of this data may have restrictions due to the need to protect patients' identities or claimants' IP, sharing it is currently impractical. Indeed, for every

---

<sup>5</sup> If a repository is required, Data.Gov.Hk can fulfil some of these needs as there are now examples of institutions such as Lincoln and Bristol Universities in the UK using the same CKAN software for research data management (Orbital, 2012). Addressing the potential for open access to private sector data, the Hong Kong Science and Technology Parks Corporation (HKSTPC) is also setting up a 'Data Studio' as a forum for developing solutions with data from the government and private companies using a brokering API (HKSTP, 2017).

argument in favour of open data, there is a parallel argument against it, as summarized in the table, below:

**Table 2. Summary of Arguments in Favour of and Against Open Data**

<i>“For” Open Data</i>	<i>“Against” Open Data</i>
Enables data-driven research and text mining	Takes resources (time and effort) to share
Increases readership and citations of papers	Potential loss of revenue from intellectual property
Increases potential for collaboration and broader reach	Data security issues/increased risk of data breaches
Enables increased citizen engagement	Creates opportunities for citizen misunderstanding
Increased trust and reputation due to transparency/reduced research misconduct	Fears of displaying your “dirty laundry”
Speeds up improvements to public health/environment	May take resources from more active measures
Helps quantify research output and productivity	Potential for gaming as previous research assessment systems
Potentially tackles issues with publication bias	Still unlikely to release negative data

Many of the government departments in Hong Kong, particularly the Companies Registry, the Lands Department and the Health Authority, collect money from selling data and they may be wary of losing this revenue. Requests for information through the Code on Access to Information have shown that the Companies Registry receives \$73,000 HKD from search fees for their data (Companies Registry, 2016), while the Lands Department receives \$4.15 million HKD (Lands Department, 2015), the Lands Registry receives \$163,969 a year from search (Lands Registry, 2018), and copy charges and the Hospital Authority receives \$250,000 HKD per year (University Grants Council, 2016) from these fees.

Such competitiveness arguments are easy to debunk when looking at how much more advanced and competitive research is in countries in Europe and North America at the top of global

transparency and open data indices. The small amounts a few departments can make back in revenue are utterly dwarfed by the additional money spent on inefficient and unused government apps, the loss of competitiveness to the Hong Kong economy and the detriment to public health from data being held back from front-line medical professionals.

Notwithstanding monetary considerations, there is a natural tension between a researcher's desire to hoard data for secondary publications for career progression and society's desire to maximally leverage data for the greater good. Research funders need to decide whom they are representing. Is it a minority of senior scientists slowly releasing secondary publications based on publicly funded data they control, is it the wider research community in Hong Kong and beyond or is it their patients and citizens, whose health and wellbeing are put at risk?

The level of openness in citizen science initiatives indicate that Hong Kong institutions are at odds with the opinions of its citizens. The RGC and researchers ought to consider whether the actions of citizens should be taken as support for change, a pseudo-democratic mandate for pursuing a more open and collaborative set of policies regarding the sharing of research objects and articles.

A change towards open research requires a both top-down and bottom-up pressures. If our exploratory results are reliable, the government, public funding bodies, institutions, researchers and the public must all become aware of the benefits of open research and lobby for the adoption of open principles. Through our investigations into the monitoring of open access publishing uptake, we observed that there is no mechanism for monitoring such detail in the UGC's CDCF and that this appears to define the information collected by research institutions. Half the institutions did not respond to our requests and our request to the UGC following the Code on Access to Information was also refused. All this indicates an air of secrecy that prevents the public and researchers from gathering the knowledge they require to provide bottom-up pressure for change. The information that was ultimately released was in PDF format rather than in openly searchable formats, indicating that open principles are not adopted across the board in Hong Kong's institutions.

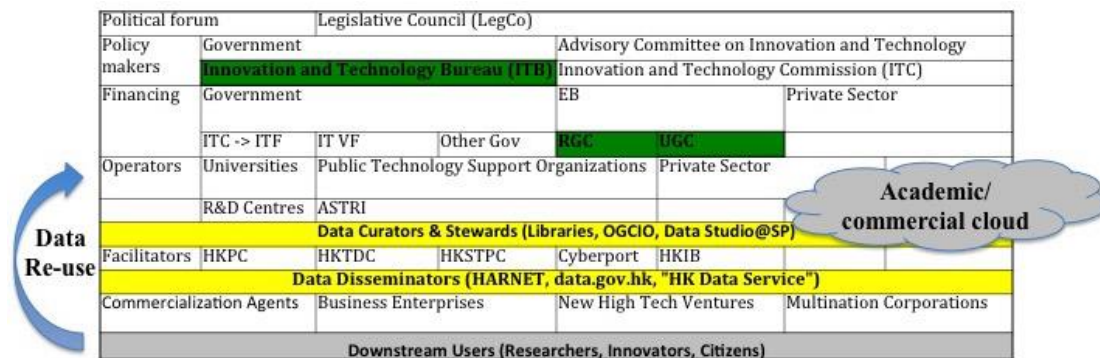
Very little has been published on this topic and the lack of transparency in even sharing statistics and information on open data and open access policies has made it particularly challenging to assess the state of play. In our investigation, we may have missed updated policies, but this puts us on the same playing field as researchers trying to figure out where they stand when publishing and releasing their research outputs.

In the latest WIPO global innovation index, Hong Kong's research inputs are very well regarded (ranking number 2 in the world due largely to Hong Kong's financial sophistication, which is

considered a strength). At the same time, research outputs are ranked much lower at rank 25 (Cornell University, INSEAD, 2016). Thus, Hong Kong’s research efficiency is very low and considered a weakness, coming in at rank 83. This sort of indexing give some indication of the loss of research efficiencies that we predict here.

Improving the conditions for sharing research data is an intricate and difficult problem as it requires understanding the ‘thick layers of complexity about the nature of data, research, innovation, and scholarship, incentives and rewards, economics and intellectual property, and public policy’ (Borgman, 2012). Yet, urgency regarding this problem seems absent in Hong Kong, as reflected in the response of the chairman of the RGC to our inquiry in October 2014 by saying that ‘the only document we can find relating to openness of research data. . . was prepared in response to a press enquiry in 2011’. The lack of leadership is demonstrated in the funding agencies stating they are merely implementation rather than policy bodies and the hierarchical players passing the buck regarding who is responsible.

We appreciate that this is complicated. An open science and open data approach would require a joint effort on the part of multiple departments to develop and institutionalise procedures and norms for a comprehensive research data policy (Leonelli & Bezuidenhout, 2015). However, Hong Kong does not need to reinvent the wheel as most other developed nations have these policies and practices either in place or on the verge of being implemented. Returning to the hierarchical representation of S&T players in Hong Kong from Figure 1, Figure 2 shows what a mature and productive data ecosystem might look like.



**Figure 2** Hierarchical Representation of Suggested S&T Data Ecosystem

Notes: OGCIO—Office of the Government Chief Information Officer

Data Studio@SP—Data Studio at the HK Science Park

HARNET— The Hong Kong Academic and Research NETwork

With a focus on maximising utility for downstream users, the ITB and university funders (highlighted in green) can demonstrate leadership through policies (mandates) and funding for new research data infrastructure (in yellow). With curators and stewards trained in data

management tasked with aiding the organisation and dissemination of funded data outputs (both academic and government data), these services can also be offered to the private sector through the HKSTPC Data Studio. A new ‘Hong Kong Data Service’, working with HARNET and government data repositories, can help to oversee these services using infrastructure from commercial clouds.

Hong Kong has a long way to go in this area. With the recently established ITB in Hong Kong, it is timely to address this issue. An open research data policy reflects a mature innovation system; technologically, Hong Kong has one of the most advanced economies, but due to a lack of awareness and the appropriate science policy mindset, the development of a sound institutional framework for research data management is lagging. Hong Kong’s new Chief Executive Carrie Lam Cheng Yuet-ngor’s maiden policy address announced a plan to double R&D spending to 1.5% of GDP (Kao, 2017). With the specific aim to boost innovation and entrepreneurship, a significant update in research policies needs to be implemented simultaneously to maximise the legacy of this significant new investment.

### **Abbreviations**

CC0: creative commons zero; CDCF: Common Data Collection Format; CPU: Central Policy Unit; CUHK: Chinese University of Hong Kong; DB: Development Bureau; FHB: Food and Health Bureau; FOI: freedom of information; HARNET: Hong Kong Academic and Research NETWORK; HKSTPC: Hong Kong Science and Technology Parks Corporation; HKU: Hong Kong University; HKUST: Hong Kong University of Science and Technology; ITB: Innovation and Technology Bureau; ITC; ITC: Innovation and Technology Commission; ITF: Innovation and Technology Fund; OECD: Organisation for Economic Cooperation and Development; OGCIO: Office of the Government Chief Information Officer; OA: Open Access; RCUK: Research Councils UK; RGC: Research Grants Council; UGC: University Grants Committee.

### **Funding**

SCE’s work was supported by BGI and China National Genebank.

### **Acknowledgements**

The authors would like to thank John Bacon-Shone, David Palmer and Fred Chan from HKU and Kostas Repanas from A\*STAR (now EMBL) for their feedback on previous versions of the manuscript.

### **Additional Supplemental Data**



The questionnaire and interviewee details are available in supplemental details available from the OSF platform <https://osf.io/4wfe9/>

## References

- AFCD. (2015). Hong Kong Live Eco-map. Retrieved June 6, 2016, from <http://www.hkecomap.net/>
- AFCD. (2018). AFCD Hong Kong Biodiversity Online. Retrieved June 6, 2016, from <http://www.afcd.gov.hk/english/conservation/hkbiodiversity/hkbiodiversity.html>
- Bobrow, M. (2015). Funders must encourage scientists to share. *Nature*, 522(7555), 129. <http://doi.org/10.1038/522129a>
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078. <http://doi.org/10.1002/asi.22634>
- Census Department. (n.d.). Census and Statistics Department homepage. Retrieved June 2, 2017, from <http://www.censtatd.gov.hk/home/index.jsp>
- China Forest Science Data Centre. (2003). China Academy of Forestry Science and Technology Office of Science and Technology letter. Retrieved December 30, 2017, from [http://www.cfsdc.org/web/module\\_html/gywm-tz2.jsp](http://www.cfsdc.org/web/module_html/gywm-tz2.jsp)
- CMA. (2017). China Meteorological Administration Data Centre policies. Retrieved December 30, 2017, from <http://data.cma.cn/DataCollect/ConcurrentPolicy/licode/4.html>
- Companies Registry. (2016). Request for information on Electronic Search Services charge statistics - a FOI request to Companies Registry. Retrieved from [https://accessinfo.hk/en/request/request\\_for\\_information\\_on\\_elect\\_2?nocache=incoming-302#incoming-302](https://accessinfo.hk/en/request/request_for_information_on_elect_2?nocache=incoming-302#incoming-302)
- Cornell University, INSEAD, and W. (2016). *The Global Innovation Index 2016: Winning with Global Innovation, Ithaca, Fontainebleau, and Geneva*. Retrieved from [http://www.wipo.int/edocs/pubdocs/en/wipo\\_pub\\_gii\\_2016.pdf](http://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2016.pdf)
- DBT India. (2014). DBT and DST Open Access Policy Policy on open access to DBT and DST funded research. Retrieved from <http://www.dbtindia.nic.in/wp-content/uploads/APPROVED-OPEN-ACCESS-POLICY-DBTDST12.12.2014.pdf>
- Earthwatch. (n.d.). Earthwatch Institute Hong Kong. Retrieved June 6, 2016, from <http://www.earthwatch.org.hk/>
- Environment Bureau. (2017). Accessinfo.hk FOI request: EB research funding and policies

- information - a FOI request to Environment Bureau. Retrieved June 11, 2017, from [https://accessinfo.hk/en/request/eb\\_research\\_funding\\_and\\_policies](https://accessinfo.hk/en/request/eb_research_funding_and_policies)
- ERAC Secretariat. (2016). *ERAC Opinion on Open Research Data*. Retrieved from <http://data.consilium.europa.eu/doc/document/ST-1202-2016-INIT/en/pdf>
- Fang, M., Yao, X., Chan, C.-K., Lau, N. T., & Lau, A. P. S. (2009). Street level air quality data collected by the Mobile Real-time Air Monitoring platform. *DataSpace@HKUST*. Retrieved from <http://repository.ust.hk/ir/Record/1783.1-3631>
- Food and Health Bureau. (2017). FHD research funding and policies information - a FOI request to Food and Health Bureau. Retrieved June 11, 2017, from [https://accessinfo.hk/en/request/fhd\\_research\\_funding\\_and\\_policie#incoming-491](https://accessinfo.hk/en/request/fhd_research_funding_and_policie#incoming-491)
- Geodata.cn. (2017). National Earth System Science Data Sharing Infrastructure - About Us. Retrieved December 30, 2017, from <http://www.geodata.cn/aboutus.html>
- Gibb, M. (2017). CityU Scholars: A hub of research excellence. Retrieved December 30, 2017, from <http://newscentre.cityu.edu.hk/media/news/2017/06/30/cityu-scholars-hub-research-excellence>
- Harnad, S. (2005). Australia is not maximising the return on its research investment. In *Proceedings National Scholarly Communications Forum*. Retrieved from <http://www.ecs.soton.ac.uk/~harnad/Temp/research-australia.doc>
- HKBWS. (2011). HKBWS Ringed BFS Record System — Ringed BFS Sightings. Retrieved June 6, 2016, from <http://bfs.hkbws.org.hk/en-US/sightings>
- HKBWS. (2014). Hong Kong Birdwatching Society - Hong Kong Bird Report. Retrieved May 31, 2016, from [http://www.hkbws.org.hk/web/eng/bird\\_report\\_eng.htm](http://www.hkbws.org.hk/web/eng/bird_report_eng.htm)
- HKSAR Government. (2016). Hong Kong Code on Access to Information. Retrieved May 31, 2016, from <http://www.access.gov.hk/en/home/index.html>
- HKSTP. (2017). Data Studio@SP. Retrieved from <http://datastudio.hkstp.org/>
- HKU Research Data Services. (2016). Research Data and Records Management. Retrieved from <http://www.rss.hku.hk/integrity/research-data-records-management>
- Holdren, J. P. (2013). *Increasing Access to the Results of Federally Funded Scientific Research*. Retrieved from [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

- Innovation and Technology Commission. (2017). Guide to Filling in the Application Forms Innovation and Technology Commission Innovation and Technology Fund Innovation and Technology Support Programme Guide to Filling in the Application Forms. Retrieved September 27, 2016, from <http://www.legco.gov.hk/yr10-11/english/panels/ci/papers/ci1116cb1>
- Kao, E. (2017). Boosting R&D in Hong Kong requires more than just funds, scientists say. *South China Morning Post*. Retrieved from <http://www.scmp.com/news/hong-kong/education/article/2116378/boosting-rd-hong-kong-requires-more-just-funds-scientists>
- Lands Department. (2015). Accessinfo.hk FOI request: Information on costs, charges and revenue for map data. Retrieved from [https://accessinfo.hk/en/request/information\\_on\\_costs\\_charges\\_and](https://accessinfo.hk/en/request/information_on_costs_charges_and)
- Lands Registry. (2018). Accessinfo.hk FOI request: Data access charges for land registry data - a FOI request to Land Registry. Retrieved August 24, 2018, from [https://accessinfo.hk/en/request/data\\_access\\_charges\\_for\\_land\\_reg#incoming-784](https://accessinfo.hk/en/request/data_access_charges_for_land_reg#incoming-784)
- Leonelli, S., & Bezuidenhout, L. (2015). Impact of Social Sciences – The Politics of Data: The rising prominence of a data-centric approach to scientific research. Retrieved from <http://blogs.lse.ac.uk/impactofsocialsciences/2015/10/05/politics-of-data-scientific-data-centrism/>
- Leyden, A. (2012). Hong Kong Air Pollution on the App Store. Retrieved June 6, 2016, from <https://itunes.apple.com/us/app/hong-kong-air-pollution/id504536152?mt=8>
- Liang, H., & Fu, K. (2015). Data from: Testing Propositions Derived from Twitter Studies: Generalization and Replication in Computational Social Science. Retrieved from <http://hub.hku.hk/cris/dataset/dataset93131>
- MOST. (2018). Opinions on Further Strengthening the Construction of Scientific Research Integrity - 中共中央办公厅 国务院办公厅印发《关于进一步加强科研诚信建设的若干意见》. Retrieved June 17, 2018, from [http://most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2018/201805/t20180531\\_139731.htm](http://most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2018/201805/t20180531_139731.htm)
- OECD. (2015). *Making Open Science a Reality*. Retrieved from <https://www.innovationpolicyplatform.org/content/open-science>
- Office of the Chief Scientist. (2014). STEM Australia's Future. Retrieved February 11, 2015, from <http://www.chiefscientist.gov.au/wp->

content/uploads/STEM\_AustraliasFuture\_Sept2014\_Web.pdf

- Orbital. (2012). Choosing CKAN for research data management. Retrieved June 6, 2016, from <http://orbital.blogs.lincoln.ac.uk/2012/09/06/choosing-ckan-for-research-data-management/>
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. <http://doi.org/10.7717/peerj.175>
- PolyU Department of Applied Physics. (2018). Community Weather Information Network (CoWIN). Retrieved June 6, 2016, from <http://weather.ap.polyu.edu.hk/>
- RDM Delivery Group. (2013). Research Data Oxford » University of Oxford Policy on the management of research data and records. Retrieved May 31, 2016, from <http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/>
- RGC. (2011). RGC HK Annual Report 2011. Retrieved May 31, 2016, from [http://www.ugc.edu.hk/big5/doc/rgc/publication/annual/rgc\\_ar2011\\_e\\_c.pdf](http://www.ugc.edu.hk/big5/doc/rgc/publication/annual/rgc_ar2011_e_c.pdf)
- ROARmap. (2014a, December 15). Agency for Science, Technology & Research (A\*STAR) ROARMAP policy page. Retrieved from <http://roarmap.eprints.org/44/>
- ROARmap. (2014b, December 15). National Chung Hsing University ROARmaps policy page. Retrieved from <http://roarmap.eprints.org/47/>
- Royal Society. (2012). *Science as an open enterprise*. Retrieved from <http://royalsociety.org/policy/projects/science-public-enterprise/report/>
- Safecast. (2016). Safecast API. Retrieved June 6, 2016, from <https://api.safecast.org/en-US>
- Shearer, K., Repanas, K., Yamaji, K., & COAR Community of Practice. (2017). *Asian Open Access: Regional Survey*. Retrieved from <https://www.coar-repositories.org/files/Asia-OA-survey-results-June-8-2017-2.pdf>
- Stringer, R. (2014). Predicting 'Red Tide' risks: Citizen Science Leader data published in Hong Kong scientific paper | Earthwatch Freshwater Watch. Retrieved June 6, 2016, from <https://freshwaterwatch.thewaterhub.org/news/predicting-red-tide-risks-citizen-science-leader-data-published-hong-kong-scientific-paper>
- Taiwan Forestry Bureau. (2016). Taiwan Forestry Department Data Policy. Retrieved May 31, 2016, from <http://www.forest.gov.tw/public/Data/64417173671.pdf>
- Taiwan MOST. (2014). Taiwan Ocean Data Bank repository Data Policy v7. Retrieved May 31, 2016, from <http://www.odt.ntu.edu.tw/wordpress/wp-content/uploads/2014/03/Data-Policy->

\_v7.pdf

Taiwan MOST. (2018). Taiwan MOST 「科技部全球資訊網. Retrieved May 31, 2016, from <https://www.most.gov.tw/>

The White House. (2017). Making Federal Research Results Available to All | whitehouse.gov. Retrieved June 1, 2017, from <https://obamawhitehouse.archives.gov/blog/2017/01/09/making-federal-research-results-available-all>

UNESCO. (2015). UNESCO Science Report: Towards 2030. Regional overview : Asia and the Pacific. Retrieved June 11, 2017, from [https://en.unesco.org/unesco\\_science\\_report/asia](https://en.unesco.org/unesco_science_report/asia)

University Grants Council. (2015). Annual review of Common Data Collection Format (CDCF) - a FOI request to UGC Secretariat. Retrieved May 31, 2016, from [https://accessinfo.hk/en/request/annual\\_review\\_of\\_common\\_data\\_col#incoming-190](https://accessinfo.hk/en/request/annual_review_of_common_data_col#incoming-190)

University Grants Council. (2016). Accessinfo.hk FOI request: Request for statistics on data charges and revenue. Retrieved from [https://accessinfo.hk/en/request/request\\_for\\_statistics\\_on\\_data\\_c](https://accessinfo.hk/en/request/request_for_statistics_on_data_c)

Van Noorden, R. (2014). Chinese agencies announce open-access policies. *Nature*. <http://doi.org/10.1038/nature.2014.15255>

Wellcome Trust DBT India Alliance. (2016). Wellcome DBT - Award Policies. Retrieved June 11, 2017, from <http://www.wellcomedbt.org/award-policies>

WWF Hong Kong. (2014). Coastal Watch. Retrieved August 24, 2018, from <https://coastalwatch.hk/home-en/>