

December 2014

**FACTSHEET**  
***Data as-a-Service Pilot***

Singapore is building to be world's first Smart Nation, with fuller use of technology to live, work and play, resulting in improved quality of life for individuals; business opportunities for enterprises and an anticipatory government that uses tech to better serve and anticipate citizens' needs. Data as-a-Service will form a key role and basic building block for Singapore's Smart Nation vision that will be underpinned by Data and Analytics.

**Overview**

The ubiquitous nature of the Internet has resulted in a massive increase in the number of digital users and activity. This increase is, in turn, generating an unprecedented amount of data. The ability to easily capture and store this data has led to it playing an increasingly significant economic role and being used to drive key decisions in many companies.

Corporations now rely on the use of high quality data for analytics to reveal opportunities for cost reduction, optimise investments, and to increase revenue. At the same time, data owners across sectors are beginning to view their datasets as not only fundamentally valuable, but also economically viable to distribute, resulting in a growing marketplace for data.

Despite the relative ease of amassing, sorting, tagging data, there are some challenges such as the difficulty in discovering datasets.

**About the Data as-a-Service (DaaS) Pilot**

The DaaS pilot seeks to study the feasibility of addressing the issue of dataset discovery in the private sector through a federated approach. Currently, there is no coherent mechanism for users to easily discover private sector datasets made available from data providers across various industry sectors.

The DaaS platform, also named the Federated Dataset Registry (FDSR), is intended to enable a mechanism to ease dataset discovery. As a federated platform, it is collectively made up of individual Dataset Registries (DSRs) based on the open-source data portal platform CKAN (Comprehensive Knowledge Archive Network) (<http://ckan.org>)

IDA encourages data providers from all industries to participate in the DaaS pilot to increase visibility of their datasets and contribute feedback on how the platform can be improved. The DaaS pilot is free to participate for data providers. For data providers who are already monetising their datasets, it would further assist them to reach out to new customers.

To participate in the DaaS pilot, interested data providers can download the DSR software by following the steps on <http://DaaS.sg>. The current list of participating data providers and the nature of the datasets (as of 10 December 2014) is appended below:

Data Provider	Nature of datasets
DFRC Pte Ltd (was BIMAR Pte Ltd)	People activity and mobility (footfall)
DP Information Network Pte Ltd	Business-related (e.g., ACRA)
DHI Water & Environment (S) Pte Ltd	Water and environment
Elixir Technology Pte Ltd	Business-related (e.g., ACRA)
Hutbitat Pte Ltd	Real estate
Jones Lang LaSalle Property Consultants Pte Ltd	Real estate Market trends and outlook
Quantum Inventions Pte Ltd	Traffic-related
Roland Turner	Air quality readings
Sense Infosys Pte Ltd	Maritime logistics
Singapore Post Ltd	Postal address
SISV Services Pte Ltd	Property listing transaction caveats
Standices Pte Ltd	Social media
StreetSine Singapore Pte Ltd	Real estate
Ventesworks Pte Ltd	Social media

The DaaS pilot that will last till 31 March 2016 will enable IDA to gain valuable feedback on the various aspects of a DaaS platform, with a focus on a new mechanism for data users to discover private sector datasets.

#### *Hosting Partner under DaaS pilot*

Amazon Web Services (AWS) has signed a Memorandum of Intent to provide up to USD\$3,000 worth of usage credits to the first 25 data providers who wish to host their DSR instance on the AWS cloud during the pilot. The credits are available for redemption over a period of 12 months. Additionally, AWS is providing up to four technical workshops and two business mentoring workshops annually for data providers.

### *Features of DaaS pilot*

Participating data providers will operate their own instance of DSR which runs on their own infrastructure (e.g. web server, cloud). Each DSR comes with a dataset catalogue which the data provider maintains. It contains details of datasets, such as meta-information and sample datasets. In addition, data providers are encouraged to profile their datasets using a set of data quality metrics, and provide these metrics for users through the dataset catalogue. This will assist users to appreciate the quality of the datasets of interest, and also allow for comparison between similar datasets. These metrics can be generated either by using a set of guidelines or with data quality tools.

Each DSR is fronted by a web portal, which participating data providers are encouraged to hyper-link from their corporate websites to provide visibility. Through a DSR instance from any participating data providers, users can explore and learn about datasets from all other participating data providers, without prior knowledge of them or their datasets.

When discovering datasets, users can filter dataset results by parameters such as data provider, dataset format, and Singapore Standard Industrial Classification (SSIC) under which datasets are categorised. Upon discovering datasets of interest, users will be able to reach out to the data providers and facilitate access to the datasets — either by monetary transfers, licensing regimes or other forms mutually agreed upon by both parties.

### *Data Quality Metrics*

The objective of the Data Quality Metrics (DQM) guidelines is to recommend a set of common metrics which is industry-domain agnostic, for adoption by data providers. The guidelines describe the methodology for deriving this set of metrics and consider tools for relaying metrics to end-users.

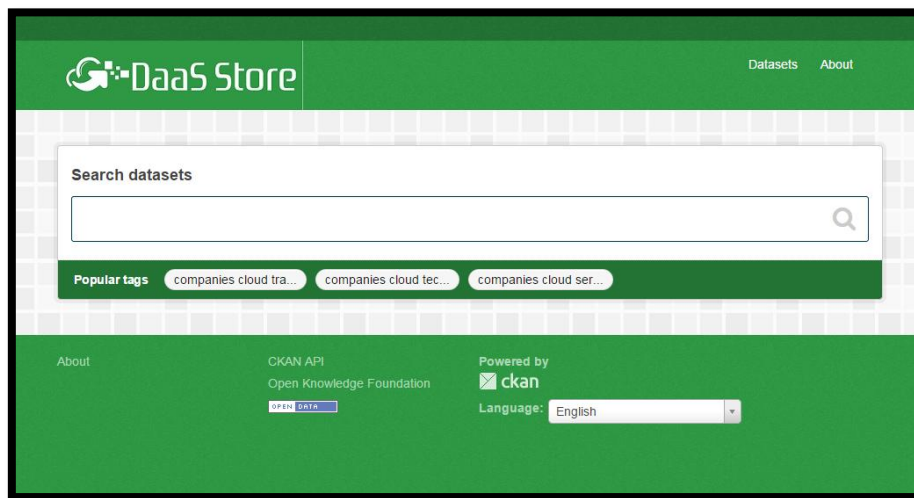
With DQM, data providers can understand and check the relative quality of their datasets in areas such as reliability, relevance, accessibility, timeliness and ease of use. For data users, it would allow them to easily compare the quality of different datasets, and match their expectations against available datasets.

Data providers can generate DQM through the use of open-source data quality tools such as the [Human Inference's DataCleaner](#), which data providers are encouraged to make use of during the DaaS pilot. To do so, data providers can simply download

the data quality tool, and follow the configuration instructions found on <http://DaaS.sg>.

A copy of the guidelines which provides detailed description of each of the quality metrics listed under these categories of quality attributes/goals is available at: <http://bit.ly/1zln2FJ>.

## Screenshots

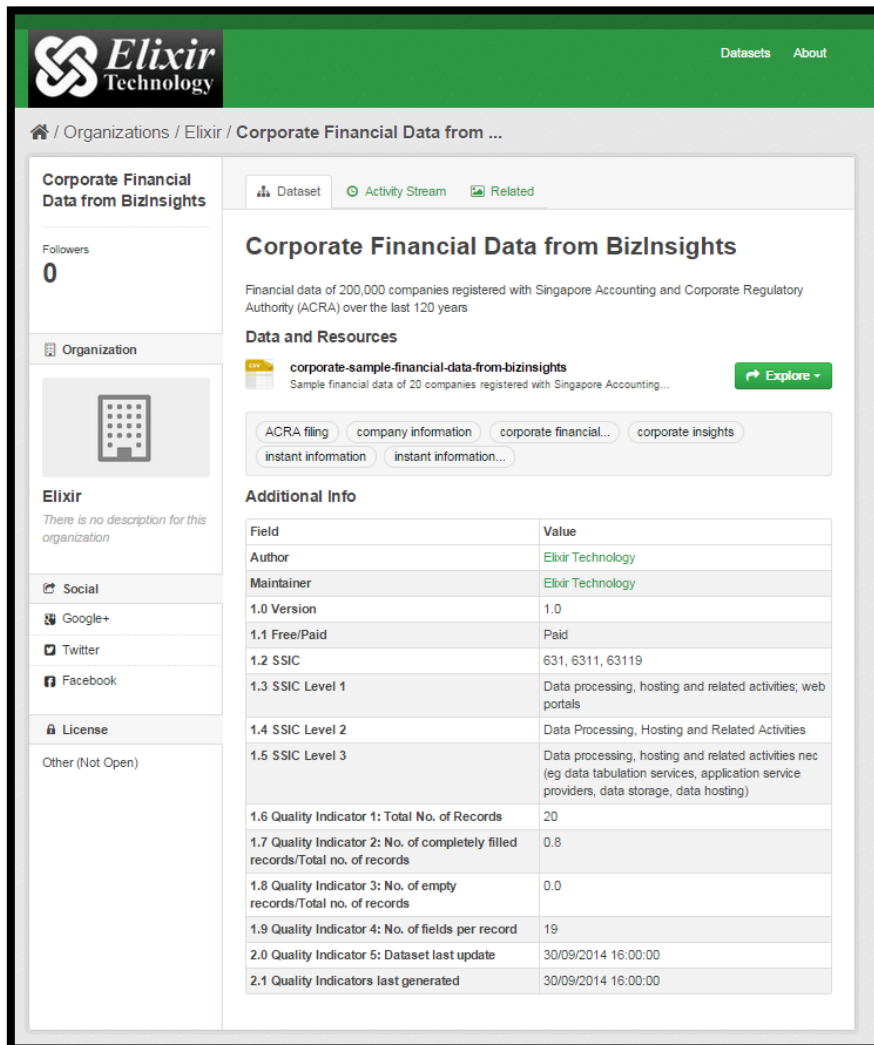


*Figure 1: Landing page of a DSR instance*

The screenshot shows the DaaS Store interface. At the top, there's a green header with the 'DaaS Store' logo and 'Datasets' and 'About' buttons. Below the header, the page title is '/ Datasets'. A search bar contains the text 'postal' and has a search icon and a 'Reset' button. Underneath is an 'Advanced Filter Options...' section with several dropdown menus: 'Data Provider' (filter by Data Provider), 'License' (filter by License), 'Format' (filter by Format), 'Tags' (filter by Tag), 'Free/Paid' (filter by Free/Paid), 'SSIC Level 1' (filter by SSIC Level 1), and 'SSIC Level 2' (Please select SSIC Level 2). Below the filters, it says '4 datasets found from 2 data providers'. The results are listed as follows:

- Query-based Real Estate Data** (Paid) - Customized query-specific data subset. From: Streetsine. File format: XLSX.
- Real Estate Listing Data** (Paid) - Exclusive listings, Co-brokered Listings, and Advertisement Listings from online portals and newspapers. From: Streetsine. File format: XLSX.
- Real Estate Transaction Data** (Paid) - Only source of Sale/Rental Transactions from Estate Agents including private sale transactions, private rental transactions, HDB sale transactions, HDB rental transactions. Transaction data from public sources is also integrated. From: Streetsine. File format: XLSX.
- SingPost 6D Postal Code** (Paid) - Postal code database. From: Singpost. File format: CSV.

Figure 2: Example of dataset search results



The screenshot shows the Elixir Technology interface for a dataset titled 'Corporate Financial Data from BizInsights'. The page includes a sidebar with navigation options like 'Followers' (0), 'Organization', 'Social' (Google+, Twitter, Facebook), and 'License'. The main content area features a description of the dataset, a 'Data and Resources' section with an 'Explore' button, and an 'Additional Info' table.

Field	Value
Author	Elixir Technology
Maintainer	Elixir Technology
1.0 Version	1.0
1.1 Free/Paid	Paid
1.2 SSIC	631, 6311, 63119
1.3 SSIC Level 1	Data processing, hosting and related activities; web portals
1.4 SSIC Level 2	Data Processing, Hosting and Related Activities
1.5 SSIC Level 3	Data processing, hosting and related activities nec (eg data tabulation services, application service providers, data storage, data hosting)
1.6 Quality Indicator 1: Total No. of Records	20
1.7 Quality Indicator 2: No. of completely filled records/Total no. of records	0.8
1.8 Quality Indicator 3: No. of empty records/Total no. of records	0.0
1.9 Quality Indicator 4: No. of fields per record	19
2.0 Quality Indicator 5: Dataset last update	30/09/2014 16:00:00
2.1 Quality Indicators last generated	30/09/2014 16:00:00

Figure 3: Example of dataset details page

**For media clarification, please contact:**

**Infocomm Development Authority of Singapore**

Eugene NEUBRONNER (Mr)

Assistant Manager, Corporate and Marketing Communication

Tel: +65 6211 1182

E-mail: [eugene\\_neubronner@ida.gov.sg](mailto:eugene_neubronner@ida.gov.sg)