USMART

HM Government
G-Cloud
Supplier

Making Data Smart

Data Sharing & Analytics Platform

G-Cloud 10 - Service Definition Document, May 2018

UrbanTide

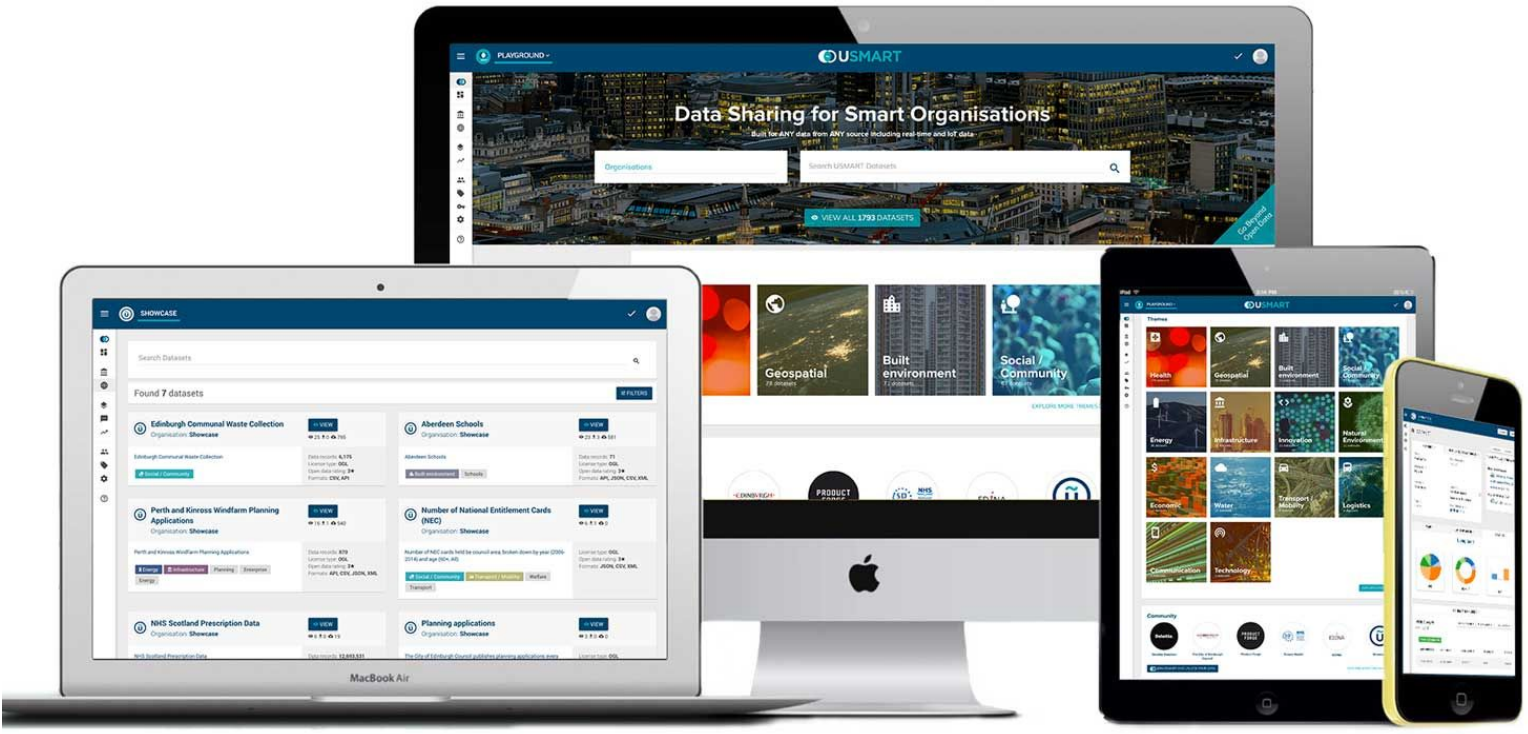
UrbanTide profile

We bring people, data and technology together to create smarter, more liveable cities for everyone.

UrbanTide formed from the team that bid, planned and delivered Innovate UK's £24m future cities demonstrator for Glasgow; the key focus was the £7.2m OPEN Glasgow programme that explored the value in new digital technologies, opened up city data and created engaging solutions for a range of urban challenges.

Since then, UrbanTide has grown from strength to strength, working with a wide variety of clients and partners. We believe we can make our cities smart by focusing on the integration of **people, data and technology** to create more liveable cities for everyone.

USMART key features



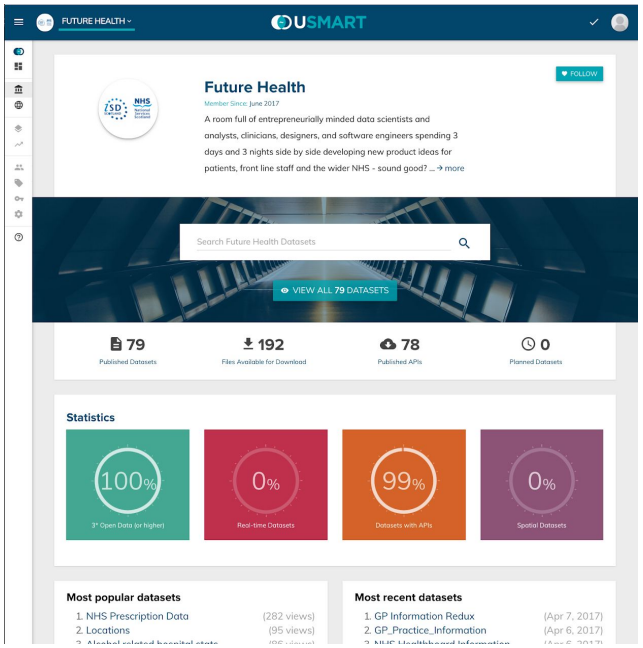
Introduction to USMART

USMART is a secure, next generation data platform that enables smart organisations to share, integrate, analyse and open up their data:

- Runs in the cloud so you can start publishing today.
- Super simple to use with minimal staff training required.
- Supports big data analytics and real-time data.
- Built to consume ANY data from ANY source.
- Open and private sharing features
- Data monetisation available.
- Complies with leading data standards (BSI PAS 182 & 183).
- Industry strength security to protect your data and ready for GDPR.

Easy data discovery

We believe it is crucial that your data is **easily discoverable and accessible**. Depending on the permissions set by the data provider, access to data can be **completely open or very restricted**. If a dataset is open, it will be made available on our global discovery pages to ensure you reach the widest audience possible.



Personalised organisation page

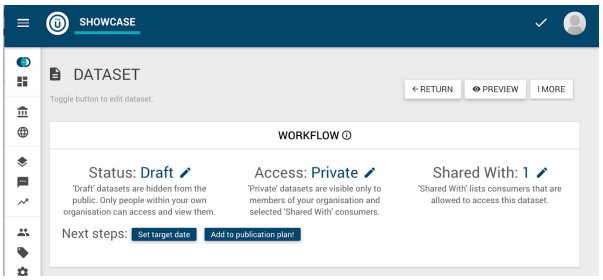
Your personalised organisation landing page for data discovery. This can be easily embedded into your own website, simplifying the need to administer another Content Management System.

USMART's social features include: saving favourite datasets to the user dashboard, commenting on datasets, contacting the dataset owner with feedback or requesting a new dataset to be published.

ANY data, ANY format, ANY size

With USMART you can consume and share any data of any format and any size. This includes big data streams and popular machine-readable formats such as CSV, XML and JSON, proprietary formats such as Excel, Word and PDF and also image formats such as JPG, PNG and TIFF. We distinguish three types of data:

- **Reference data** that doesn't change very often.
- **Batch data** where new data is added in regular intervals (weekly/monthly).
- **Real-time data** generated by back office systems and/or sensors that constantly send updates.



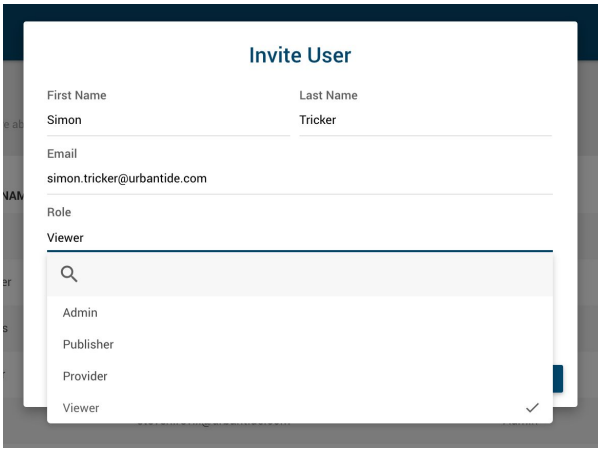
Super simple data management

We help you to complete all the required metadata fields with our **unique workflow wizard**, ensuring that data entry is consistent and validated. This also includes a three-step process of data publication from **'draft'** and **'planned'** stages to **'published'**.

USMART supports three main levels of sharing:

1. **Closed:** only members of the organisation have access.
2. **Private sharing:** members of the organisation plus selected other individuals have access.
3. **Public sharing:** everyone has access.

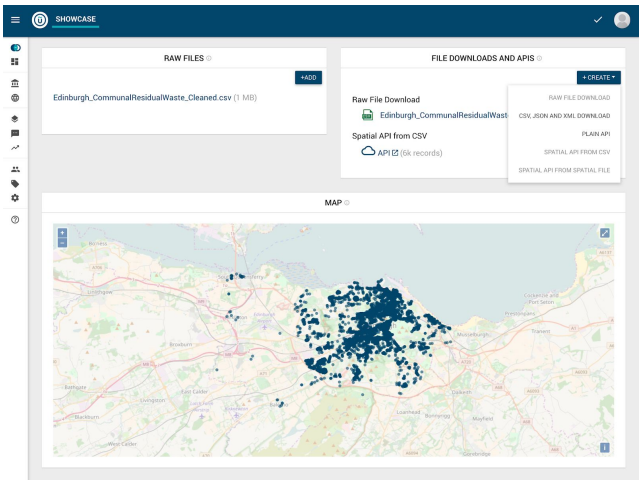
These access levels can be set on a per dataset basis and therefore provide you with full control over how closed or open your data is.



User permission levels

Access to USMART features is controlled via fine-grained user permission levels:

Public Consumer: has access to personal social features and open data.
Private Consumer: has access to privately shared datasets.
Viewer: all Consumer features plus can view all organisation datasets, including 'draft', 'planned' and 'published' states.
Provider: all Viewer features plus can create new and edit existing datasets.
Publisher: all Provider features plus can change the state of datasets to 'published' and can also unpublish datasets.
Admin: all Publisher features plus can add/ remove user accounts and set user permissions.

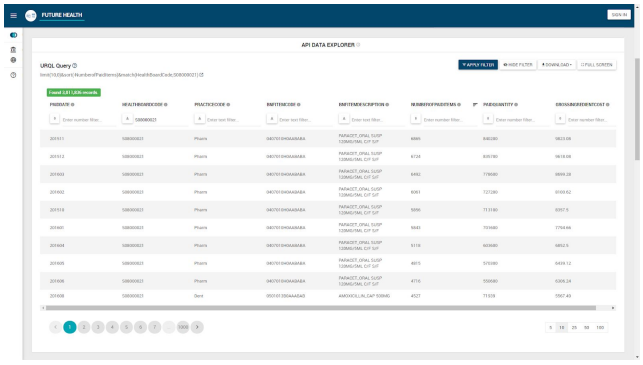


Data pipelines and data transformations

USMART innovative data pipelines can transform your raw data. **For example, a CSV file can be transformed into XML and JSON and published in these new formats in addition to the original format.** If a dataset is comprised of multiple files (e.g. weekly/monthly air quality measurements), USMART can combine these raw files into a single file or API (application programming interface) output, if required. **This is one of the key features of USMART and we'll be adding new data pipelines based on customer feedback.**

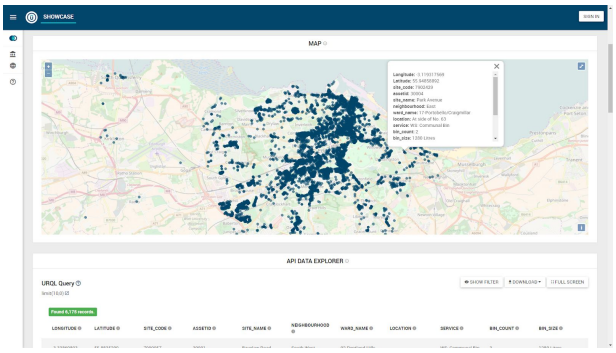
Visualisations, analytics and data stories

USMART features a comprehensive suite of interactive and easy-to-use tools for data visualisation and analytics.



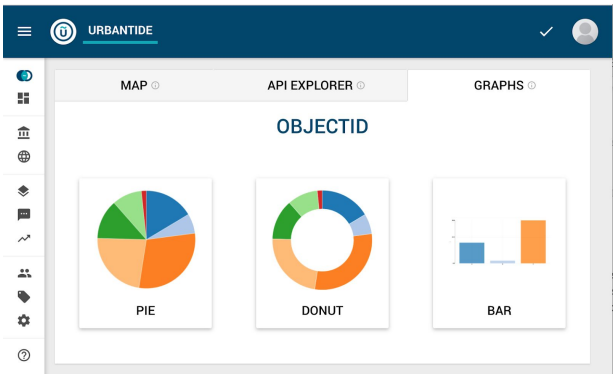
Excel on steroids

For tabular data USMART automatically generates a Data Explorer that allows users to sort, search and filter the dataset. Filtering can be done on keywords and also by unique values.
USMART's Data Explorer tool can handle **millions of data rows without impacting on performance** and it features an integrated download option allowing the user to extract and download a subset of the data, **for example**, *only the data for one or several specified wards.*



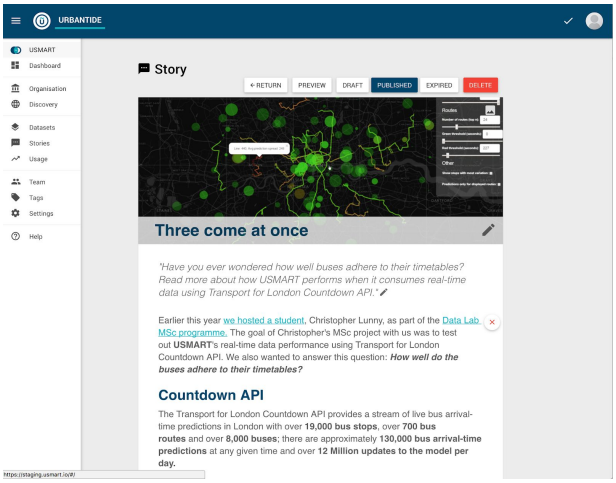
Automatic map generation

For spatial data, the dataset details page will **automatically** display an interactive map that can be used to explore the spatial data. The user can click on each of the map features to retrieve and display detailed information in a pop-up window. USMART can load and visualise **ALL** types of spatial data **(points/lines/polygons)** from many spatial data source formats **(ESRI Shapefile, MapInfo Table, GML, KML, etc)**. Our platform will also generate spatial datasets from non-spatial CSV files if they contain point coordinates.



Graph tools

We are currently implementing a new feature allowing any user to generate graphs against any tabular dataset. This feature will include bar, line and pie charts and users will be able to save charts onto their dashboard and share them with other users.



Data stories

Data stories are a great way to engage with the public and help to bring life to raw data and provide context. To support this, USMART features our integrated stories module that can be used to create articles and blogs about datasets. These stories can also be featured on your organisation's home page.

Making data smart with big data and real-time data

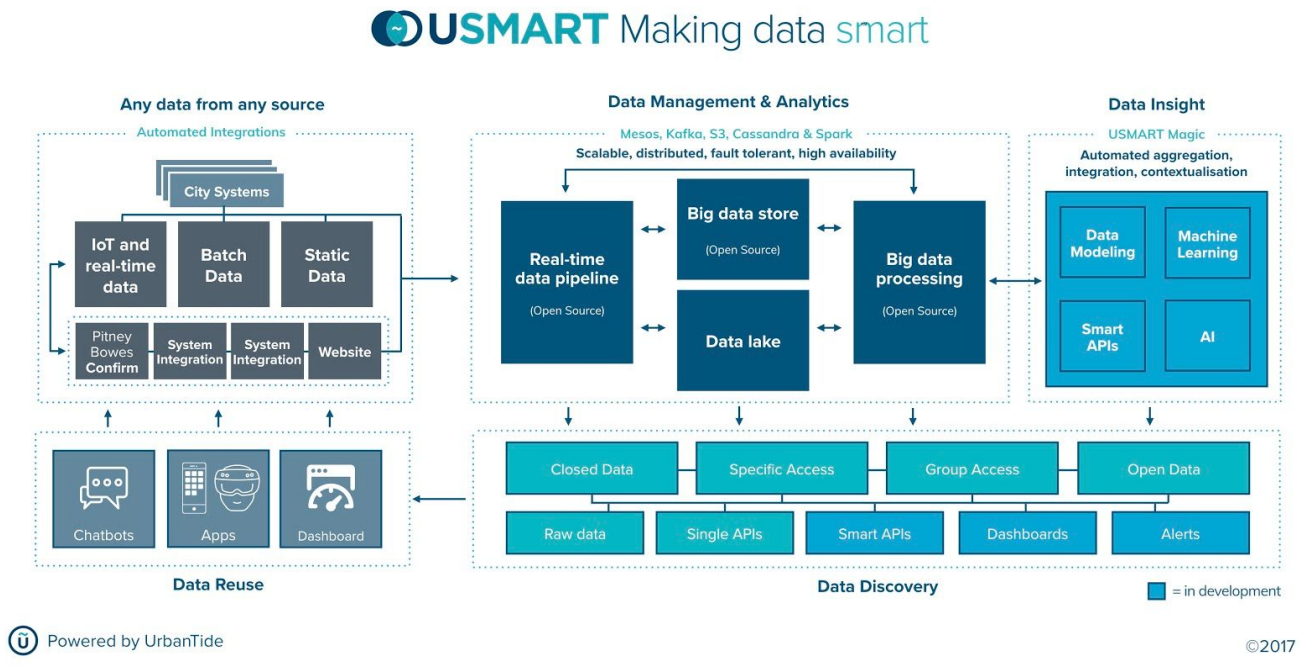
We have developed USMART using the latest big data and cloud technologies. This allows our clients to consume and publish real-time data streams using our unified, high-throughput and low-latency platform.

To publish real-time data in USMART you must create a real-time dataset. You will then be able to programmatically push data to the dataset using our API (REST or websocket). Authentication is required for this action and can be granted via an API key generated in USMART.

Data consumers can subscribe to a dataset's data stream using HTTP streams and websocket connections. Whenever you push a data packet to the datasets, each subscribed client will automatically receive this.

Since our real-time infrastructure is using the same scalable technologies that underpin all of our other microservices, we can easily scale the streaming service based on demand.

For example if you are publishing a real-time traffic feed, our system will ensure that the required computing resources are made available to meet this demand.



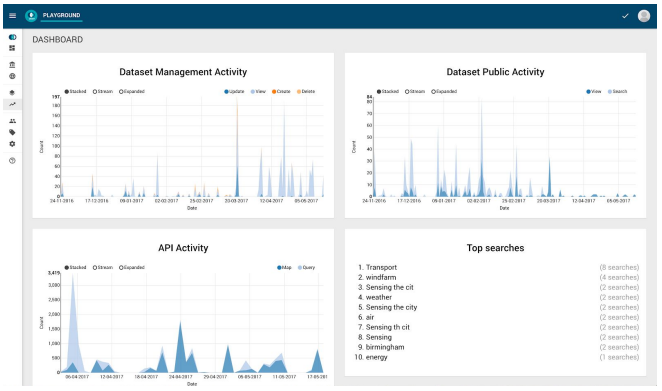
Big data analytics

USMART is a data science ready platform. Real-time data stored within our system can be analysed using Apache Spark, one of the leading open source big data processing solutions.[1]

Your data scientist can use programming languages such as Python or Scala to execute data processing algorithms that generate new insights from your data. Processed data can be shared and accessed in USMART just like any other dataset and secure permission controls apply.

Big data storage is provided by Apache Cassandra, a fully scalable time-series optimised data store.[2]

Consumer insight



USMART automatically tracks data consumer activity and provides usage statistics and graphs to system managers. We also log searches performed via our data discovery functionality, providing useful insights into search patterns and areas of user interest.

Data harvesting and federation

In January 2018 we re-launched our data harvesting endpoint that will allow other data discovery portals, including data.gov.uk and existing web pages, to index your USMART datasets and make them accessible for searching in line with the common approach in data.gov.uk via the federated model. The endpoint provides the metadata in the DCAT application profile and is fully compliant with the necessary standards.

Monetisation

We have developed the data monetisation component to enable organisations to sell their data via USMART. If you would like to find out more about these features please get in touch as we would love to beta test the functionality with you.

[1] https://spark.apache.org/
[2] http://cassandra.apache.org/

Technical details

Scalable architecture

We are using a server architecture that is inherently scalable, self-healing and supports real-time data ingestion and processing. Our data and metadata storage facility is fully elastic and there are no fixed limits. We achieve this scalability due to our modern microservice architecture. This allows us to scale each of our services to match the demand generated on it.

Deployment

USMART is available in the cloud via the Internet and can be accessed from within your internal network. This Software-as-a-Service (SaaS) model has great advantages as it doesn't require any software to be installed by your organisation. Our user interface is fully responsive and works across desktop PCs, laptop PCs and mobile devices running iOS, Android, Windows or Linux. We support one of the last two versions of leading browsers: Internet Explorer, Firefox, Chrome and Safari.

Availability

We have built our USMART platform using the latest cloud technologies which ensures high availability, redundancy and resilience. In the last 30 days we had uptime across all our services of 99.992%, far surpassing a standard 99.8% uptime requirement. Our platform is built using microservices that can be individually scaled based on demand and to ensure resilience.

Updates

Because USMART runs in the cloud, you will always benefit from the latest version with the latest features. We fix major issue immediately and release hotfixes and minor releases every two weeks. This means that any issues will be resolved quickly and without any additional cost incurred by you. Major new features are generally released every 4-8 weeks and clients are

welcome to preview and test new features on our staging site before they reach general availability.

Security

All data exchange with USMART is fully encrypted, ensuring that your data is kept safe. Data encryption at rest is also available. User access is tightly controlled and fine-grained permission roles define the available system actions for each user. User credentials must comply with our strict password policies. All actions carried out on USMART are tracked and stored in our integrated audit system.

Interoperability

USMART can publish data in various different forms. If the raw data is of a proprietary nature (e.g. Word/Excel/PDF), the system will only be able to publish the original file format. If the raw data is CSV, our platform can publish the original format, transform the data into XML/JSON and publish a powerful data API (JSON). If the raw data is a spatial file (e.g. ESRI Shapefile, MapInfo Table, GML, etc) the system can publish the original file and a spatial API (GeoJSON).Our data APIs are documented openly on Github and are used internally by our own USMART data explorer.

Virus scanning

When files are exchanged in USMART, our integrated virus scanning service will automatically scan the files for threats and block the transaction if an infected file is discovered.

Recoverability

USMART complies with a Recovery Point Objective (RPO) of 6 business hours. Our backup strategy takes full snapshots every night plus we are able to restore data from the commit log to any point in time over the last 7 days.

Web standards

USMART fully supports HTML 5.0, CSS 3 and the standard character set ISO/IEC 8859-1 (Latin1). Our platform is compatible with WCAG 2.0 AA accessibility and the eGIF interoperability framework v6.1.

Many thanks and please get in touch if you require any further information.

Steven Revill, Co-founder & CEO

Telephone: 07921 912422
Email: steven.revill@urbantide.com

UrbanTide, Codebase, 38 Castle Terrace, Edinburgh, EH3 9DZ