

White Paper

Navigating Open Data Standards for App Creation: The Citadel Story



#OpenData
@Citadel_eu

Authored by:

Julia Glidden, Senior Associate Fellow at the Institute for European Studies, VUB

Geert Mareels, eGovernment Manager, Flemish eGovernment Authority

Florian Daniel, Developer and Consultant, Items International

Susie Ruston McAleer, Managing Partner, 21c Consultancy

Ben Cave, Head of Research, 21c Consultancy

Navigating Open Data Standards

Citadel on the Move is grounded in the conviction that the power of Open Data remains untapped. There are a number of implementation projects that focus on technically advanced cities with active Open Data communities. There is also a growing body of Open Data standards work that likewise reflects the requirements and needs of more technically advanced stakeholders. To date, however, little work has been done to ensure that Open Data tools and standards reflect the on the ground reality of the vast majority of cities around the world - many of which are under-resourced and struggling to come to terms with the fundamentals of opening let alone using public data.

Citadel on the Move set forth to help narrow this gap by targeting cities and citizens with little or no open data experience and enabling anyone, regardless of their technical skills, to create an app from a basic spreadsheet in a matter of minutes. The project did so by creating a number of tools that made it relatively easy to convert simple Point of Interest data formats into richer ones more suitable for app development and then testing these new tools with a pragmatic community of users across Europe and around the world.

Throughout the course of this testing process it became clear that Citadel had unwittingly unleashed a 'data quality through publication' process wherein data users (civil servants and citizens) helped data producers (cities) to identify a number of recurring issues which often made it difficult if not impossible to make meaningful use of government data. These issues reflect 'bottom up' insights into the way in which existing Open Data Standards are currently functioning in the real world.

This White Paper focuses on the lessons that Citadel on the Move has learned in the Open Data arena with the aim of identifying the way in which project findings can help in the advancement of outstanding standards needs. It begins by reviewing the tools that Citadel on the Move has created to make it easier to open and use data held by governments as well as any other organization, association, business or citizen in a city. It then examines the 'Citadel Model' or 'usage paths' and processes from data conversion and publication through to the creation of the mobile apps before discussing the existing standards that Citadel on the Move considered and used, and the rationale therein. Finally, the paper examines the gaps in existing standards that our work with over 120 cities across all 6 continents encountered and provides recommendations for further data standards work.

Citadel Tools

Citizen Resources

To deliver on the project's twin ambitions of raising awareness about Open Data and empowering people to use it, Citadel developed a number of tools to support the publication and reuse of Open Data. These tools are hosted on the Citadel website - citadelonthemove.eu – and are specifically designed to help users create simple mobile applications:

- **Data Converter** - The converter transforms source datasets, which primarily take the form of simple spreadsheets or 'tabular data', i.e. CSV (Comma-Separated Values) or Excel files, to a specifically-structured JSON (Java Script Object Notation) file. The resulting enriched format, called 'Citadel JSON', basically tells the mobile applications

how they should display the dataset through a process called “semantic mapping” which lets a mobile app access a clear meaning for every column in the original dataset.

- **Open Data Index** – The index catalogues datasets held in Citadel JSON format either in our datastore or elsewhere on the Internet. Any dataset referenced in this index can be used by Citadel mobile applications.
- **Application Generation Tool (AGT)** – The AGT enables any person to quickly and easily build mobile applications or ‘apps’ by selecting the desired dataset(s) from the Open Data Index and personalising the appearance of the resulting app. More tech-savvy users can access several ‘Application Templates,’ each designed to handle specific data types including displaying places, events, or parking places availability.
- **App Catalogue** – The catalogue lists all the apps that are created on the Citadel platform.
- **Dataset Authoring Tool** – This tool enables local administrations without any Open Data skills to manually create a dataset that is automatically converted to Citadel JSON.
- **Dataset Testing Service** – This service enables a user test a dataset to make sure it is compatible with Citadel’s AGT.

In addition to the above tools, the Citadel website contains a wide array of documentation and support including a discussion forum, a Q&A section, video instructions and reference documents with links to Github for more advanced technical users.

Developer Resources

Although Citadel is primarily focused on empowering ordinary citizens to use Open Data, the project also elected to use Github - the first worldwide platform used by developers to host collaborative projects¹ - to offer its tools to the open source community and enable developers to use, improve and complement them. The resulting Citadel on the Move Github space - github.com/CitadelOnTheMove enables teams of developers to collaborate on the project and includes:

- **Documentation Pages** – These pages explain how to get started and link to the detailed documentation pages and tutorials for each tool
- **Converter Tools** - These tools can be used to convert various data formats to Citadel format, or from Citadel format to other formats
- **Mobile Application Templates** – These templates are ‘ready-to-use’ mobile applications that can be used as a starting point for ‘citizen developers’ with a degree of technical knowledge to create custom apps
- **Application Generator Tool (AGT)** – This tool and its code is available for organisations that would like to host their own mobile application generation platform.

Though Citadel’s Github presence is aimed at a more technology-literate audience, the tools can be used both by Local administrators to provide their own Open Data infrastructure to foster mobile application creation locally and by software developers to further enrich the Citadel toolkit in the future.

¹More than 7 million users working on more than 16,3 million projects in October 2014

Solution Overview - The Citadel Model

Beyond providing public access points for Citadel tools, the project placed great importance on developing a technical model that links key resources or datasets in the context of an Open Data journey. The resulting Citadel Model below represents the datasets as Input and Output data formats that are processed by Citadel Middleware (Citadel tools) and finally used to create Applications (mobile apps):

Dataset	Data Store	Processor	App. Template
Data that is available under a specific format, as a file or as a web ressource.	Stores or indexes where the data or dataset files are made available.	Tools that are used to generate, convert, enrich or transform the data.	Mobile applications that illustrate use cases, and can be adapted to own need (template).



Actor	Direct / automatic	Indirect / manual
Organization or person who performs an action.	Direct relation between two elements = does not require manual intervention after initial configuration.	Indirect relation = requires a manual intervention each time it is triggered / processed.

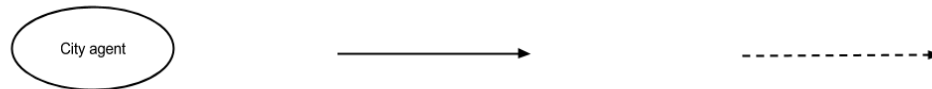


Figure 1: Citadel Data Schema

The Citadel model enables a user to visualize processes that combine several tools and understand how they can be linked to produce a mobile application from Open Data:

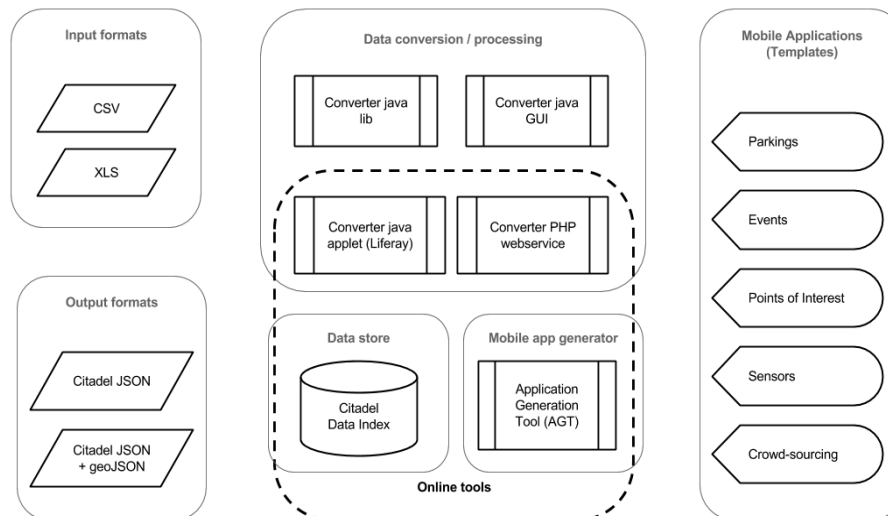


Figure 2: Citadel Tools Schema

Citadel Solution Walkthrough - Creating an App

The following steps outline the way in which anyone can use the Citadel solution to open convert, publish and build a mobile application:

1. At the beginning of the process, the data has to be available in the format most commonly used by local government - as a spreadsheet, either in **XLS** or **CSV**² - and must contain the minimal mandatory fields, which are a title and geographic coordinates³.
2. The 'source' file is next converted to the Citadel JSON format (discussed in further detail below) using the 'Converter'. The conversion process performs 'semantic mapping' which links columns headings to specific categories and then adds necessary information about the dataset (such as its author and license) to enable mobile applications to know how to interpret the data and display it on a map.
3. The converted file is then published to the Citadel Open Data Index.
4. The Application Generator Tool next lets the user select the dataset from the Open Data Index and create a mobile application in a matter of seconds.
5. The resulting application can be accessed through a dedicated URL, emailed to friends, or downloaded to any smartphone.

The following table outlines the Citadel app creation process:

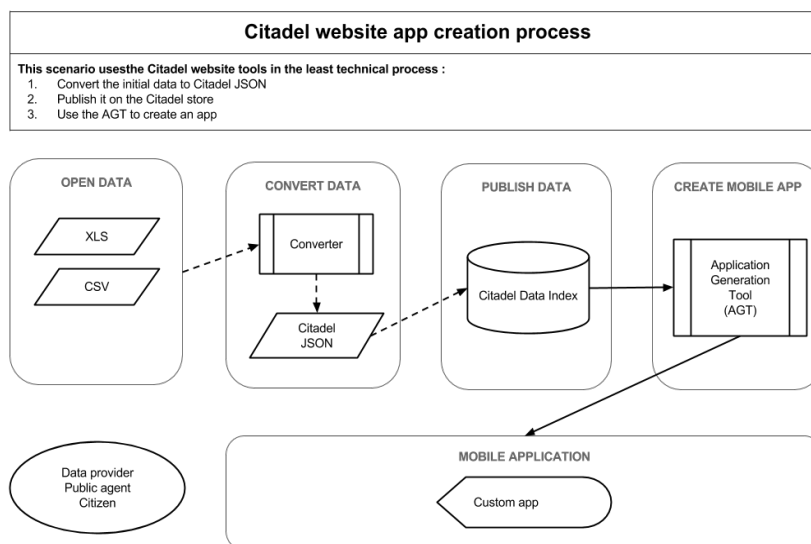


Figure 3: Citadel App Creation Process

Solution Walkthrough – Further Capabilities

² CSV files, or Comma Separated Values files are text files that contain table data. CSV files are used by 62% of the cities in the Citadel Associate network and preferred by many Open Data managers to other common data formats including XLS, XML and RDF for their simplicity and use. In addition, CSV has superior interoperability with a wider range of data transformation and enrichment tools. Most software that handles "table data" is able to save it in the CSV format, including well-known Excel, Google Docs, and databases.

³ The input data has to include geographic information, which usually means having latitude and longitude fields. These can be obtained by converting a "text" address (street, city, etc.) to geographical coordinates, which is called "geocoding." Numerous online tools offer geocoding services.

The Citadel App Creation Process outlined above is the most common way that users interact with Citadel tools. Citadel's commitment to open source code through our Github presence also enables any developer to adapt and improve our tools, and foster their reuse. The Citadel repository⁴ further accelerates this process through 'quick-start' documentation for developers⁵. One example of the developer community's work with Citadel has been a PHP version of the converter⁶ which enables webservices that consume 'live' data sources in Citadel apps. In addition to development, Open Source connectors and converters between Citadel and other projects have also been built with MyNeighborhood⁷, CitySDK⁸ and FI-Ware⁹ projects.

Beyond this process, however, the Citadel toolbox also enables users to explore a much wider range of scenarios including:

- Tweaking application templates to personal specifications and branding (All Templates)
- Collecting user-generated POI data (Crowd-Sourcing Template)
- Integrating data feeds from real-time sensors (Environmental Data Template)
- Using web-services to convert files on request (Citadel PHP converter)
- Updating data in an app real-time (Live Dataset URL from Database Backend)
- Hosting one's own AGT and Convertor (Github Source Code)

The Citadel Focus – Driving Use Across Europe

Engaging Local Administrations

Citadel tools have been built with the active participation and feedback of our pilot cities of Athens, Ghent, Issy-les-Moulineaux and Manchester and our Associate Network which, as of December 2014, includes **120+ Local Administrations** from **46 European countries** and across **all 6 continents**. Collectively, these efforts have generated the following results:

- **640** Mobile Apps built using the Application Generator Tool¹⁰
- **356** Datasets published or referenced in the Citadel "Open Data Index", in 13 categories¹¹
- **18** Local Authorities opening their first dataset with Citadel
- **44** 1 on 1 consultations with City administrators
- **63** Countries Worldwide with at least 1 Citadel Associate
- **19** Associates in UK – the most for a single country

Assessment of Open Data Practices

As part of the Citadel Associate Outreach Programme, the Citadel team assessed the Open Data activities and practices of Associates in 3 central areas:

⁴<https://github.com/CitadelOnTheMove>

⁵<http://citadelonthemove.github.io/>

⁶<https://github.com/CitadelOnTheMove/converter-php-lib>

⁷ [http:// myneighbourhood.eu](http://myneighbourhood.eu)

⁸ [http:// citysdk.eu](http://citysdk.eu)

⁹ [http:// fi-ware.org](http://fi-ware.org)

¹⁰<http://www.citadelonthemove.eu/en-us/exploremycitadel/appcatalogue.aspx>

¹¹<http://www.citadelonthemove.eu/en-us/opendata/opendataindex.aspx>

1. Local Administration's Level of Open Data Maturity

The first area focused on the level of 'sophistication' of Associates' Open Data efforts:

- **17%** had no previous contact with Open Data (no data was public available for the associate).
- **24%** had little experience of Open Data (some data but no city portal or systematic release).
- **47%** had some experience of Open Data (a city portal or systematic release but no clear policy on Open Data publication & updates)
- **12%** had advanced experience of Open Data (a portal or systematic release and a policy of Open Data publication & updates)

2. Local Administration's Open Data Formats

The second area focused on the formats in which Open Data was published. The format of the Open Data is here defined as the file structure of the Open Data rather than the specific information it contains. Our assessment found the following:

- **77** distinct formats were used to share Open Data.
- **62%** of administrations use CSV (Comma Separated Values) to publish at least one of their Open Data files
- **2-3** different formats was the average used per administration

The following figure shows the percentage of associates using the 10 most popular formats:

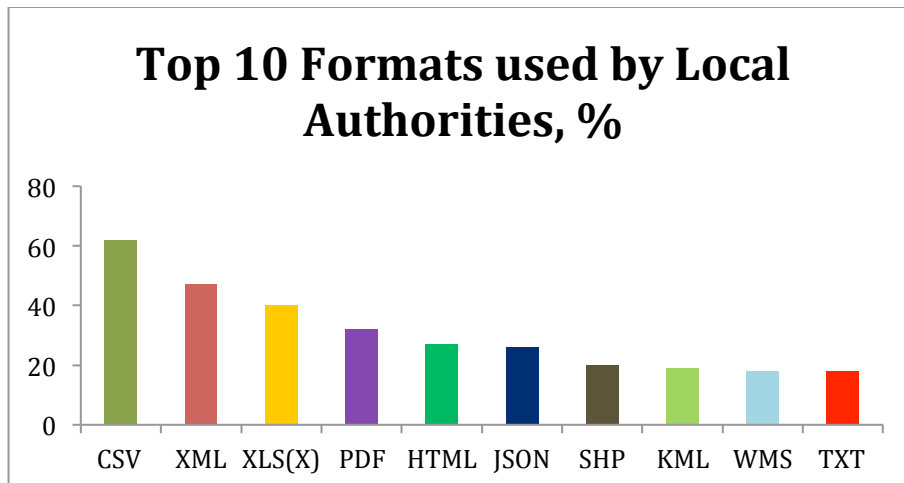


Figure 4: Citadel Associates Data Format Choices

3. Type of Portal Used to Publish Data

The final area concerned the use of Open Data portals. Our team found that:

- **10%** of Associates had no publically-accessible Open Data portal
- **17%** used a National Open Data portal
- **44%** used a Local Open Data portal

- 29% used both Local and National portals

Citadel's findings on our associates Open Data practices highlight one of the key findings of the project: **A significant gap exists between 'Standards' (as set by global technical bodies) and the common practices which cities are actually using.**

Citadel Approach to Standards, Formats & Licenses

Approach to Standards

Standards can often be confusing and contradictory, particularly in the still relatively new Open Data domain. As a project devoted to increasing the practical usage of Open Data, Citadel therefore consciously chose to adopt a pragmatic approach to Open Data standards based on the following principles:

1. Use existing open standards and formats when available
2. Rely on well-known and widely-used technologies
3. Prioritise interoperability with as many systems and organisational processes as possible

Approach to Formats

As an excellent analysis of data standards by Tim Davies¹² argues, in the early years of Open Data, many standards advocates, including Citadel in its initial White Paper, championed Linked Open Data (an approach that represents data as a series of interconnected links) because its way of representing data, primarily using RDF, can be used to build very advanced models and tools. Through time, however, many came to reject the LOD model because, as our own work on Citadel discovered, it is highly complex to create and work with and therefore excludes less technically advanced data owners. For a short while, it then began to look as though XML (a format created by W3C) represented a better choice for publishing Open Data because it contains a strong 'schema' to organize datasets, is easily available to export data from standard city web portals, and provides easy retrieval to applications consuming the data. However, there has likewise been a move away from the XML standard over the past two years due to the difficulty of building the rigid schema into apps.¹³

More recently, the Open Data community has begun to embrace the notion that simpler 'tabular' data formats offer the best prospect for reuse by the widest possible community, including those for whom the more technically complex formats like XML represent a barrier to entry. In light of this trend toward 'flat' schema-free data, CSV (Comma Separated Values) has gained popularity in recent years as the best general-purpose format for releasing Open Data. W3C's Technical Architecture Group is currently working on a draft to provide guidelines on CSV syntax and best practices and has even gone as far as to declare 2014 'The Year of CSV.'

¹² <http://www.opendataimpacts.net/2014/10/data-standards-and-inclusion-in-the-network-society/>

¹³ *ibid*

Citadel's work with local authorities across Europe and subsequent findings on data formats supports the W3C position. Whilst certain sections of the technical community, notably those committed to furthering the 'Web of Data' vision of Berners-Lee¹⁴, may still advocate the advanced capabilities of LOD (represented through RDF), Citadel's findings show that this 'ideal' ignores the reality on the ground in four senses:

1. **City Data Owners** overwhelmingly prefer the simplicity of CSV as a publication format
2. **Open Data Advocates** promote CSV for 'data dumps' as this format provides greater clarity than APIs or RDF¹⁵
3. A growing number of **Developers** prefer CSV tables as this format represent a 'blank canvas' on which they can work more effectively¹⁶
4. **Ordinary Citizens** wishing to use data can understand CSV files without the advanced technical skills required for RDF or XML.

Indeed, the trend toward flat datasets like CSV suggests to us that the less provenance or 'baggage' a dataset carries the more useful it is for all users regardless of technical ability¹⁷.

Licensing

Dataset licenses are extremely important because they define how, and under what conditions, Open Data can be reused. Citadel recommends in all cases that a City choose the 'most open' license possible to encourage maximum data reuse, minimize barriers to development and ultimately create a vibrant community of data re-users. As licensing depends on various legal requirements, Citadel focused on selecting a set of licenses appropriate for Open Data, rather than on a single license. After considering a range of license sets including GPL (which has a republication clause that may limit some commercial reuse) and BSD (which is less onerous but still places some requirements on reuse) the project settled on the 'Creative Commons'¹⁸ set of licenses. Citadel recommends using 'Creative Commons' licenses because they are very popular with Open Data providers, widely accepted and simple to understand.

Citadel specifically recommends using CC-0 'No rights reserved' wherever possible. CC-0 is similar to placing the data 'in the public domain', allowing any use of the data without restriction. A CC-BY license (which preserves attribution - meaning a data user has to credit the source of the data) is recommended to our associates for any conditions where CC-0 cannot be applied for legal reasons. Citadel does not recommend other licenses from the CC family such as CC-ND (which prevents the distribution of 'altered' forms of the data) or CC-NC (which prohibits commercial re-use) as these do not conform to open principles. The chosen license should also be explicitly defined and freely available for the user to see.

¹⁴ <http://www.w3.org/2013/data/>

¹⁵ <http://theodi.org/blog/2014-the-year-of-csv>

¹⁶ <http://data.okfn.org/doc/csv>

¹⁷ A more extensive discussion of this point can be found in Citadel deliverable D3.3.2 'Semnatic Framework for Data Integration'

¹⁸ www.creativecommons.org

Citadel Use of Formats & Standards

File Formats Mapping

From our interactions with Pilot and Associate cities, Citadel quickly understood that most data owners in cities do not have a strong grasp of the relative benefits of different data models and file formats. Therefore, all but the most advanced cities take the path of least resistance; publishing data in whatever data structure they already hold using spreadsheet-friendly file formats as .CSV and .XLS. In only a few cases (10%) did Associate cities have any data published in an RDF-model format.

In light of this on-the-ground reality, the Citadel team realized that it needed to create tools and associated recommendations that balanced the simple compliance steps required to get cities on board on the one hand, with maximizing data reuse on the other. Toward this end, the team conducted a mapping of the most appropriate file formats available to ascertain their pros and cons:

Citadel Common File Formats Mapping Grid

Format	Description	Pros	Cons
.XLS/.XLSX (Excel)	Represents data as tables accepted by all spreadsheet programmes	Very accessible to people and widely used	Proprietary format. Does not use Unicode. Too simple to allow for most programmes to make use of this data directly. Does not express relations between data
.CSV	Represents data in 'flat' tables easily read by people or machines	Easily understood or 'parsed' by most programmes. Easily read by humans. Application-neutral.	Tabular format does not express relationships between data making it less applicable for complex applications
.XML	Represents data as a structured tree schema that expressed relations between data	Strong schema makes it possible to attach rich information data	Schemas are complex making it difficult to write programmes
.JSON	Represents data in a simple tree structure making it programmer-friendly	Structure and links make it very easy to build services using data	Lack of schema means it only supports simpler types of data
RDF Formats (e.g Turtle,	Represents data as a network of linked points that make it easy to	Highly-structured information makes it easy to search and	Complex structure makes it difficult to 'parse' and therefore costly to work with.

N-Triples and JSON-LD)	understand complex patterns using computer programmes	retrieve information in services. Easy to visualise data relationships	Relational structure makes human reading challenging. Complex to work with in development.
------------------------	---	--	--

Based upon the above mapping and in alignment with the trends outlined above, Citadel ultimately chose to use two types of files formats in its own work out:

- CSV – Recommended for publishing Open Data for the public¹⁹,
- JSON – Recommended for publishing Open Data to use in mobile applications

File Formats Choices

Citadel’s choice to use CSV as the main input data format was first and foremost led by the project’s ambition to foster easy access to open government data. As discussed, a majority of cities around the world use spreadsheet programmes such as Excel for managing their data. Thus, rather than fight this trend and try to impose a more technically advanced standard from above, Citadel elected to work with the practice via a format like CSV that is easy to edit and export²⁰ and noted for its simplicity and accessibility²¹, compactness, ease and speed of processing and scalability²². In addition, CSV is a commonly-used format for ‘data dumps,’ which are used by the Open Data community to build databases and webservice.

The choice of JSON for mobile development was motivated primarily by a desire to make our Open Datasets attractive to App developers. JSON presents data in a simple ‘tree’ structure without the need for a formal ‘schema.’ JSON can also be read directly by applications whereas XLS and CSV cannot. JSON is consequently well-suited to building capable apps and popular with the developer community²³ because, unlike ‘schema’ formats like XML, the dataset contains all the information the app needs to work well. At the same time, JSON’s lack of schema likewise makes it relatively easy to convert standard CSV files into this developer-friendly format.

Data Model Choice

To support the release of developer-friendly JSON, Citadel created a common data model for each dataset – which we call ‘Citadel JSON’ (technical title C-JSON). Citadel JSON’s Data Model is based on an extension of the W3C POI Core Data Model.²⁴ Citadel JSON’s data model has two advantages over other commonly-used JSON data models:

1. **Semantic Annotation** – A Citadel JSON file tags every piece of information with a machine-readable category. These tags are contained within the dataset, meaning any app can easily read and make sense of the data without additional programming. The uniform structure means an application always finds the correct data in the expected location with the right label attached to it.

¹⁹ Some of the Citadel converting tools also accepts Excel, OSM JSON or geoJSON files

²⁰ <http://www.opendataimpacts.net/2014/10/data-standards-and-inclusion-in-the-network-society/>

²¹ CSV files can be read and edited by humans, using a simple text editor

²² Huge CSV datasets can be easily handled, as the “one line per entry” structure enables sequential processing

²³ <http://blog.mongolab.com/2011/03/why-is-json-so-popular-developers-want-out-of-the-syntax-business/>

²⁴ <http://www.w3.org/TR/poi-core/>

2. **Cross-Border Use** – A Citadel JSON file will work with any other application designed using Citadel JSON. This means that an app developed to find art galleries in Helsinki can also find galleries in Palermo with no need to develop and download a new service.

The two features above make Citadel JSON a significant improvement, from the perspective of developers, over existing Open Data models. The following visual provides an overview of the Citadel JSON data model for POI Datasets:

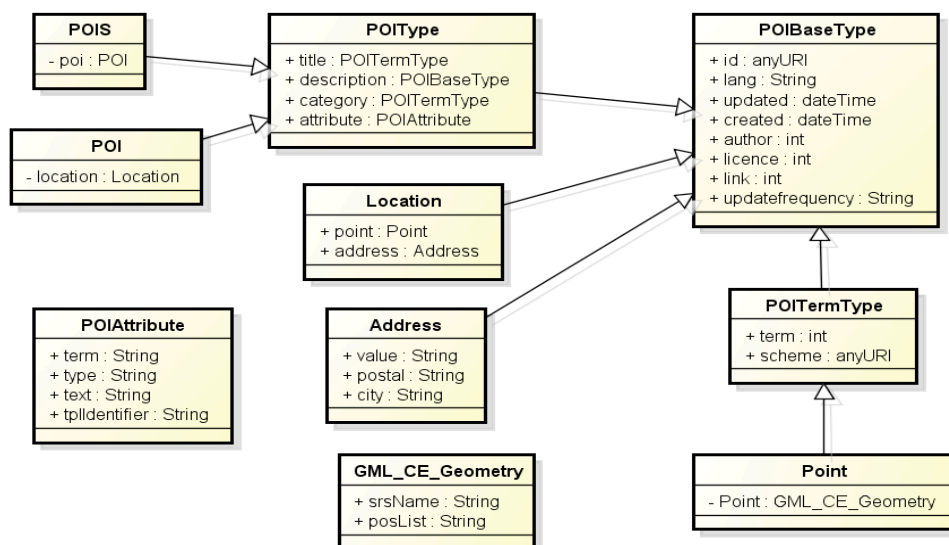


Figure 5: Citadel JSON Data Model

The Citadel JSON POI data model is a significant extension of the W3C POI Core Draft, currently the global guideline for the production of POI Data Models. In the Annex of this document, you can find a full mapping of the Citadel JSON Data Model against the POI Core Draft to understand the extensions and differences.

Points of Interest (POI) Standards

As noted above, Citadel’s Data Model is based on the W3C POI Core draft²⁵ which defines a data model for ‘location about which information is available.’ As the model was designed to be used in mobile web applications, it was implemented in JSON which, as discussed above, is the most used and suitable data format²⁶. The resulting format called “Citadel JSON” includes both data from the original dataset, information about the dataset itself (known as “metadata”: “data about data” – see below), and specific fields that describe how the data should be used into mobile applications.

Citadel JSON can be compared with the related format GeoJSON, which is also used to describe POI, but does not include information about the dataset, and is not specifically designed to provide all required data for mobile applications. Both files formats can be merged, or easily converted.

²⁵<http://www.w3.org/2010/POI/documents/Core/core-20111216.html>

²⁶JSON is the native data format used by JavaScript, which is responsible for the dynamic part of the applications

Geospatial Standards

Citadel JSON uses the WGS84 coordinate reference system to represent the location of points of interest accurately. WGS84 is used by GPS providers as well as most well-known mapping systems and can be considered the world standard. The coordinates of a given point on Earth are expressed into decimal format, using axis order latitude, then longitude, and separated with a space, e.g. “50.838908 4.373942” for the European Parliament building in Brussels. The Citadel JSON conversion process also uses latitude and longitude fields to produce the Citadel JSON file. The resulting format in Citadel JSON combines these two fields with a separating space, resulting in a single value with latitude first, then longitude.

The Citadel tools also allow other standards to be declared and converted using the Citadel Converter, though this is not recommended and not handled by the mobile templates²⁷.

Date and Time Data

Citadel uses date and time information both to identify the dataset itself and as part of datasets related to events. The Events Template created by Citadel uses this information to filter the POI based on selected dates. The dataset time metadata is formatted using the ISO 8601²⁸, preferably with time zone shift information (e.g. “2014-10-02T15:13:19+00:00”), as the applications may be used in various time zones.

Due to the variety of input data formats in source datasets, the date format used for Events POI is inputted free-form into Citadel JSON. However, precautions should be taken to ensure that the format used can be parsed by the Events Template.

In addition to ISO 8601, Citadel Events Templates notably handles RFC2822 Date and Time Specification²⁹ (eg. “Mon, 25 Dec 1995 13:30:00 GMT”), as well as dates using the format “DD/MM/YYYY”.

Sensor and IoT

As part of Citadel’s Pilot Activities with the City of Manchester, we integrated some real-time sensor data into our App Templates. Citadel work on sensor data needed to find a relevant standard and ultimately led the project to the work of the Open Geospatial Consortium (OGC)³⁰ on Sensor Web Enablement. OGC has developed a series of industry-leading data model standards³¹ that are designed to describe, query and exchange sensor information. Of these model standards, two proved useful for the Citadel sensor data:

1. “Sensor Observation Service” (SOS)³² - describes a web service to query sensor data,
2. ISO 9156:2011 “Observations and measurements”³³ - describes the sensor data.

²⁷Because the underlying map uses WGS84 too

²⁸<http://www.iso.org/iso/iso8601>

²⁹<https://www.ietf.org/rfc/rfc2822.txt>

³⁰<http://www.opengeospatial.org/>

³¹These standards are: SensorML (sensor model), O&M (observation data), SOS (observation service), SPS (planning service), and SAS (alert service)

³²<http://www.opengeospatial.org/standards/sos>

³³http://www.iso.org/iso/catalogue_detail.htm?csnumber=32574

Data produced by sensors and by Internet of Things (“IoT”) devices are often only available in proprietary formats, or vendor-specific data models. Such data feeds are generally accessed through a webservice (an automatic dataset that supplies the latest information when requested by an app or service) which most commonly export data in JSON, XML and CSV data format. Citadel chose to use the proprietary sensor platforms³⁴ already installed in Manchester to showcase live sensor data in action³⁵. For others exploring the use of sensor data as Open Data, we recommend the use of a webservice with either JSON or CSV format.

Metadata

Metadata, or ‘data about data,’ is ‘structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource.’³⁶ The reference metadata standard to describe online resources is Dublin Core Metadata,³⁷ from which 15 core terms³⁸ have been normalised in ISO 15836:2009.³⁹ Citadel chose to conform to this widely-recognised international standard - all Citadel data therefore uses Dublin-Core Metadata.

POI Dataset Categories

ISO 19115⁴⁰ describe the main themes for geographic dataset categorization. These top level thematics have been extended by the INSPIRE directive which recommends for each dataset:

- A unique INSPIRE theme⁴¹
- Additional keywords from the GEMET-Concepts,⁴² or a professional thesaurus or free keywords

The Citadel Data Index⁴³ used a more pragmatic approach that gave rise to a different, narrower categories list as it is more convenient for general public POI categorization. This classification is available using a JSON implementation of RDF Data Catalogue standard (DCAT)⁴⁴ through the dataset webservice of the Open Data Index.

The “categories” used inside the datasets themselves are free because they reflect the ones used in the original data file which may or may not be structured. While not enforced at all, Citadel’s use of existing categorization vocabularies can be a step forward toward better interoperability. It would allow better dataset auto-discovery in the future, and is therefore advised.

³⁴ The Manchester sensors use Xively, which is fast-becoming an industry standard for sensors. For more information see here: <https://xively.com/dev/>

³⁵ Which uses Xively API, which was used by 2 pilot cities, fetching data as JSON : <https://xively.com/dev/docs/api/>

³⁶ http://en.wikipedia.org/wiki/Metadata_standards

³⁷ <http://dublincore.org/>

³⁸ The 15 core terms are : title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights

³⁹ http://www.iso.org/iso/fr/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=52142

⁴⁰ http://www.iso.org/iso/fr/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53798

⁴¹ <http://inspire.ec.europa.eu/theme/>

⁴² <http://www.eionet.europa.eu/gemet/en/themes/>

⁴³ <http://www.citadelonthemove.eu/en-us/opendata/opendataindex.aspx>

⁴⁴ <http://www.w3.org/TR/vocab-dcat/>

Mobile Application Templates

Citadel's mobile applications templates use HTML5⁴⁵ – which is mobile-platform agnostic – rather than vendor-specific mobile platform technologies such as Google's Android or Apple's iOS⁴⁶. While HTML5 itself doesn't allow developers to build native apps (an app program developed for use on a particular platform or device) at this point in time,⁴⁷ it can be embedded into native applications for iOS and Android to offer native features to mobile users. Citadel's HTML-based templates can be used even on the most basic web servers,⁴⁸ and adapted without a heavy development environment, using only text editors.

Gaps in Existing Standards

Character Encoding Issues

Citadel supports the UTF-8 character encoding standard – which has existed since 1996, been a world standard since 2003 and supports all known alphabets on earth. Despite the benefits of UTF-8, use of Citadel tools has identified a number of outstanding issues regarding character encoding done using a different format:

- Some popular spreadsheet editing software including Microsoft Excel still uses regional character encodings by default
- This choice leads to less-interoperable files, with accents and special characters being misinterpreted
- Such regional encoding can cause challenges for the conversion process including unreadable accent characters in files

Citadel Recommendation:

- Adoption of the UTF-8 standard should be ensured so that text data and information can be exchanged in an interoperable manner throughout Europe and across the world.
 - Where Possible Public Authorities should say their dataset programmes to use Unicode⁴⁹
- Access to state-of-the-art standards and their implementation into varying languages should be free wherever possible.⁵⁰

POI Issues

As W3C did not finalise a standard for POI, Citadel had to use the unfinished draft. Use of this draft uncovered a number of issues:

- Some fields did not suit “flat files,”⁵¹ and were simplified to become more user-friendly,

⁴⁵HTML5 standards also rely on CSS3 and JavaScript languages

⁴⁶Despite being open-source, Android remains complex to use for non-developers, and iOS is proprietary

⁴⁷A big issue with HTML5

⁴⁸The templates also use PHP, and MySQL for database-powered apps, which both power most of the world's websites

⁴⁹<http://www.csvimproved.com/en/frequently-asked-questions/916-save-a-csv-file-as-utf-8>

⁵⁰Eg. ISO standards are still paid-access.

⁵¹They are rather designed for Linked Data, using extensive URI and namespaces instead of clear text and URL, which are more user-friendly for developers and data editors that lack the surrounding infrastructure to easily produce these structured and linked data files

- Citadel needed to describe the dataset itself, and extend POI drafts to add fields that describe the dataset – these fields basically wrap the updated POI data,
- Citadel needed to add fields specifically designed to be used with mobile applications, which were implemented through an additional extensible data model (the “tpl” identifiers),
- The draft W3c standard did not go far enough - real-world implementation revealed a need to build usable tools without using a full linked data infrastructure.

Citadel Recommendation:

- Foster the development of probed standards that fit developer needs.

Events Issues

The W3C POI data standard did not include calendar information by default which made it impossible to properly display events on the map.

Citadel Recommendation:

- Citadel added to POI data using the extendable attributes defined in the Citadel JSON format.

Geospatial Issues

Geospatial standards still are a confusing standards area. Many reference systems exist with no clear guidance on which ones are best to use and why.

In our work, Citadel found that geographical coordinates from different countries had variations in both axis order (whether latitude or longitude comes first when written) and the form in which Lat/Long were written (one cell or two separate cells) Citadel also found that geographic coordinates are based on various geospatial reference systems which often leads to inconsistency between datasets as they are not always explicit (especially once used on table files). As an example, Barcelona – which has published many rich data sets – offers a bus stop dataset, which even after proper conversion to the global standard for Latitude and Longitude (Wgs84), shows a small shift (about 2 blocks) for all POIs - making it concretely unusable for Citadel apps.

Conversion between geographic coordinate systems remains a complex issue for non-GIS specialists. We believe this complexity has contributed to a lack of easy-to-use available tools and best-practice on used geospatial reference systems. Finally, the auto-discovery feature of the Citadel AGT, which allows any app to automatically detect data corresponding to one’s current location and load it into the app, shows that there is a key need to be able to describe the covered area of a given dataset (instead of attaching it to a central point) in order to enable applications to get the most accurate data at different scales depending on their completeness. A possible pre-requisite for this functionality would be a common thesaurus of administrative boundaries and their evolution through time.

Citadel Recommendation:

- Set a central open interoperable standard for geographic coordinates based on WGS84 and a once defined axis order and coordinates formatting

- Provide adequate and open conversion tools to enable data publishers to publish their data using a shared, unique coordinate system to exchange information outside from the GIS community
- Establish an administrative ontology of European boundaries, at scales and with a historical perspective, in order to enable local data naming and discovery using both geographical coverage and administrative entities
- Where possible, cities should geocode their POI in latitude and longitude fields

Conclusion

Citadel on the Move's underlying philosophy is that standards do not develop top down from pure theory, but rather from the popularity of the applications and systems that use them. As such, the project began by implementing existing standards and then complementing them with new advances as needed.

Citadel will continue to engage more and more people in the debate around Open Data standards, using the momentum of our extensive associate community to reach beyond the technical arena and bring in those at the forefront of Open Data inside local authorities. In addition to this engagement, the Citadel team will continue to champion our pragmatic approach to standards by pushing for simple, open source tools that solve common challenges like geographic coordinates and character encoding. Finally, Citadel continues to provide a pragmatic voice within the ongoing push toward semantic standards, sharing the experiences of our associates and participating in consultations and meetings wherever possible.

Throughout the project, our goal was to create a framework within which to experiment with and probe existing standards and a space within which standards can converge. By unlocking and facilitating a lively exchange of experiences between our associates, however, we believe that Citadel has also helped to pave the way for a new and more open standardization process wherein users are able to comment upon, review and propose improvements in standards as they function in the real world rather than simply adapt themselves to that which has been imposed from above.

www.citadelonthemove.eu



Citadel on the Move



@Citadel_eu

Citadel on the Move is a 3 year project part funded by the European Commission's ICT Policy Support Programme (ICT PSP) under the Competitiveness and Innovation Programme (CIP)



Citadel on the Move is led by the Flemish eGovernment Authority (CORVE) supported by:

