

Refactoring data delivery:

The case study of the new tools for
census flow data at UKDS

Vassilis Routsis / Oliver Duke-Williams

UK Data Service and UCL Department of Information Studies

v.routsis@ucl.ac.uk



The UK Data Service and census data



UKDS provides access to the latest as well as past census data from 1971 onwards



It brings expertise, knowledge, and vast experience on handling census data. Some of our colleagues have been working on census data for decades



Separate groups within UK Data Service work with different types of census data



Flow data

Flow data enumerate movements of people from one place to another. They include migration, journeys to schools or work, and travel to second residences.

These data are characterised by their complexity, volume, and sparsity.

Data are available from censuses from 1981 onwards. Data from 2011 and 2021 have different levels of access.



Web Interface to Census Interaction Data (WICID)

WICID, initially developed in the late 1990s, went online in 2000 to provide a web-based interface for flow data. It underwent a redesign in the early 2000s to support the 2001 census outputs and received significant upgrades in the early 2010s to accommodate 2011 census data. Today, WICID remains the only web tool in the UK dedicated to flow data, offering sub-setting capabilities and hosting census data spanning four decades.



wicid WICID → Query → Data → Select by table [Select table] Help
Logged in as: guest Customise Logout now

[Census links] [WICID project] [WICID query] [Current query] [Save / restore queries]

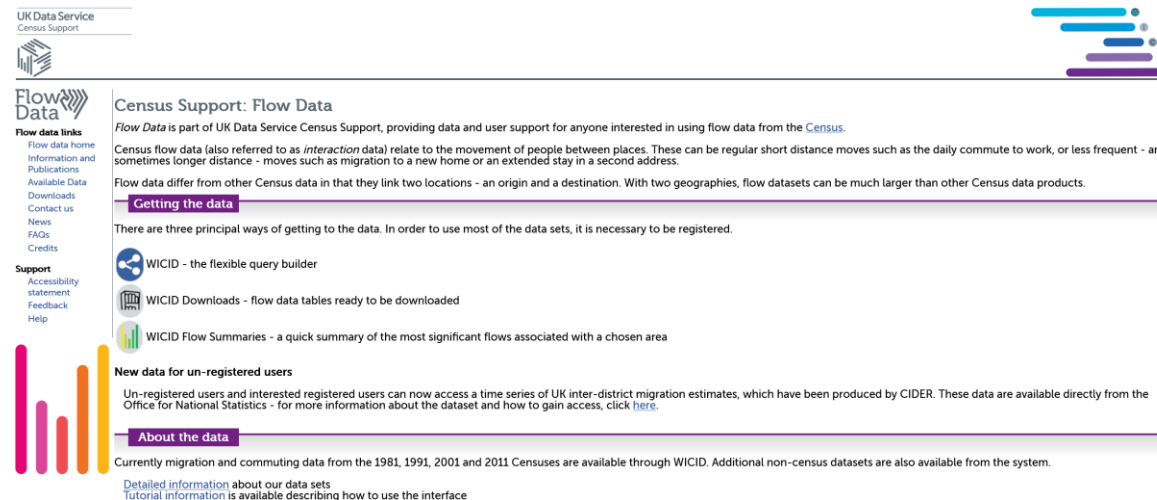
0 data items currently selected [Unselect all] [Edit list]

- Table 1 [All migrants: age \(5 broad age groups\) by sex](#)
- Table 2 [Wholly Moving Households and residents in Wholly Moving Households: counts](#)
- Table 3 [All migrants: age \(5 year groups\) by sex](#)
- Table 4 [All migrants: marital status by sex](#)
- Table 5 [All migrants: ethnic group](#)
- Table 6 [All migrants: whether resident in households by whether suffering from limiting long term illness](#)
- Table 7 [All migrants aged 16+: economic position](#)
- Table 8 [Wholly Moving Households: tenure](#)
- Table 8S [Wholly Moving Households: tenure](#)
- Table 9 [Wholly Moving Households: sex and economic position of head](#)
- Table 10 [Residents in Wholly Moving Households: sex and economic position of head](#)
- Table 11S [All migrants: Gaelic speakers](#)
- Table 11W [All migrants: Welsh speakers](#)

What now?

You must select some data before you can produce any output!

- Select one of the above tables
- OR go back to the [dataset selection](#) list to choose a different data set
- OR go back to the general [data selection](#) page
- OR go back to the general [query](#) interface



UK Data Service
Census Support

Flow Data

Census Support: Flow Data

Flow Data is part of UK Data Service Census Support, providing data and user support for anyone interested in using flow data from the Census.

Census flow data (also referred to as *interaction* data) relate to the movement of people between places. These can be regular short distance moves such as the daily commute to work, or less frequent - and sometimes longer distance - moves such as migration to a new home or an extended stay in a second address.

Flow data differ from other Census data in that they link two locations - an origin and a destination. With two geographies, flow datasets can be much larger than other Census data products.

Getting the data

There are three principal ways of getting to the data. In order to use most of the data sets, it is necessary to be registered.

- WICID - the flexible query builder
- WICID Downloads - flow data tables ready to be downloaded
- WICID Flow Summaries - a quick summary of the most significant flows associated with a chosen area

New data for un-registered users

Un-registered users and interested registered users can now access a time series of UK inter-district migration estimates, which have been produced by CIDER. These data are available directly from the Office for National Statistics - for more information about the dataset and how to gain access, click [here](#).

About the data

Currently migration and commuting data from the 1981, 1991, 2001 and 2011 Censuses are available through WICID. Additional non-census datasets are also available from the system.

[Detailed information about our data sets](#)
[Tutorial information](#) is available describing how to use the interface

<https://wicid.ukdataservice.ac.uk/>

New systems for census data

CKAN for aggregate data

DCAT for improved data discovery

API-driven tools for flow data to supersede WICID



Why was a new platform needed?

Modernisation and future-proof design

- A platform using latest technologies and best-practice principles designed to last for many years in the future.

Interoperability

- A dedicated fully-functional API allows the exchange of information with other systems – whether internal or external.

Flexibility, customisation and modularity

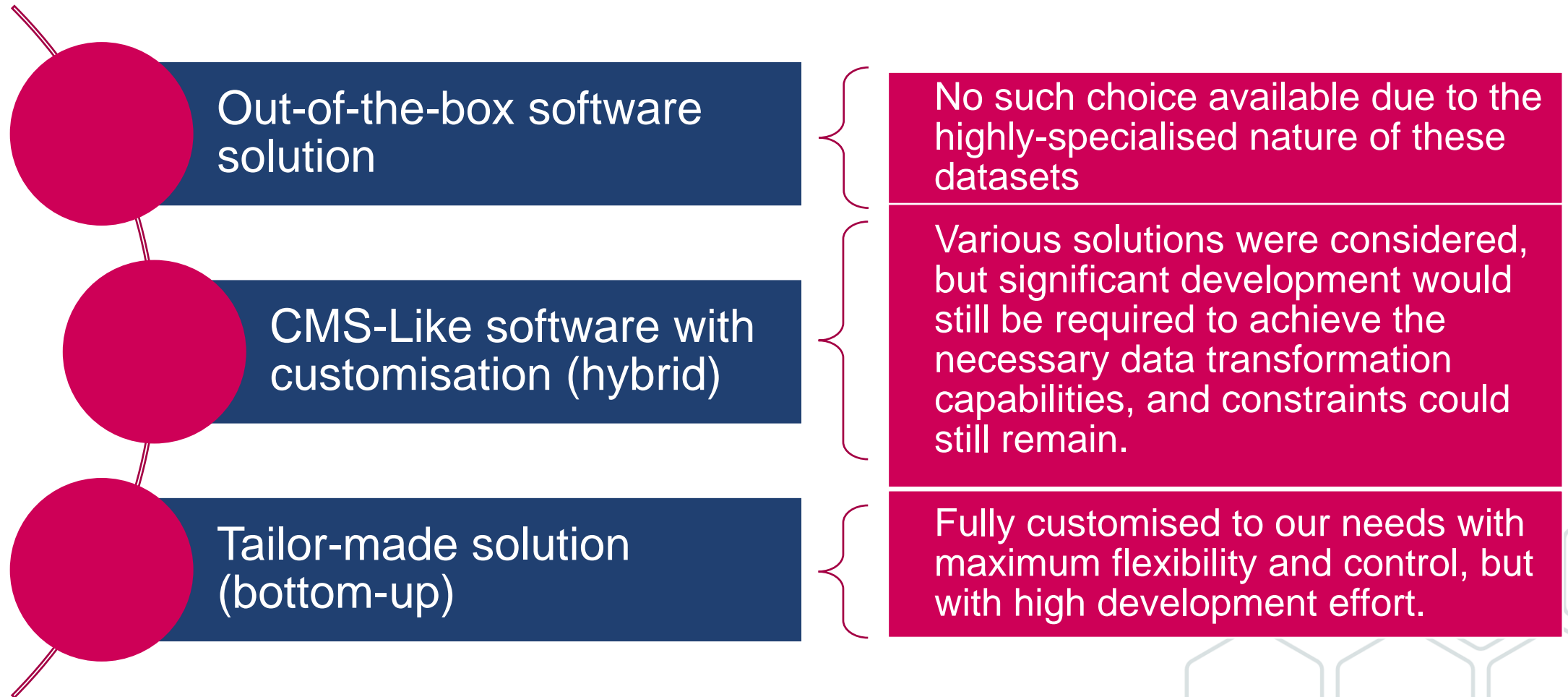
- Separation of backend and frontend allows flexibility on the interface design, quicker adaptability to changing requirements and advanced modularity as components can be customised, extended, and implemented as needed.

Developing a modern platform for the delivery of highly specialised datasets

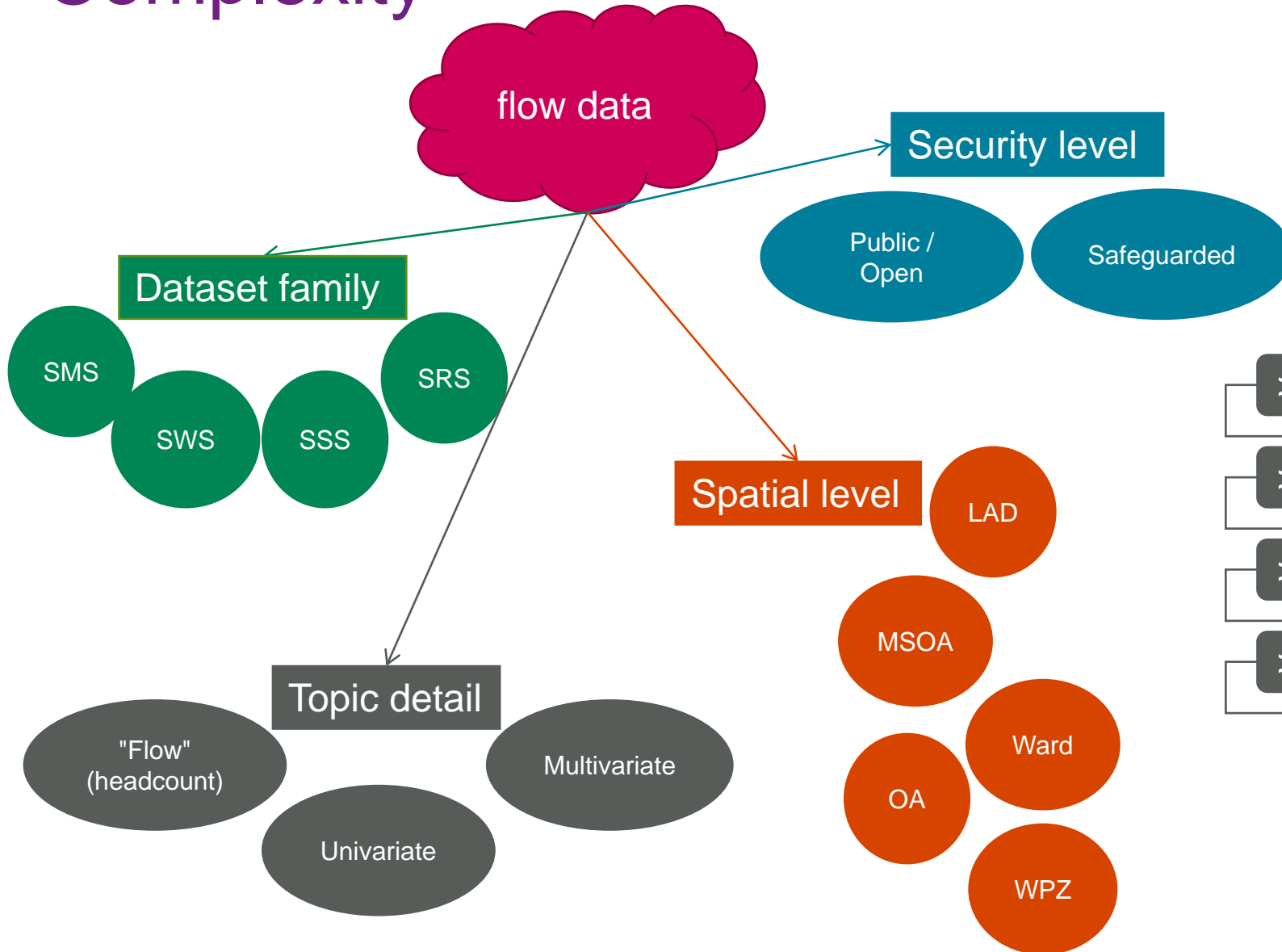


Software development conceptual decisions.

Build vs buy dilemma



Complexity

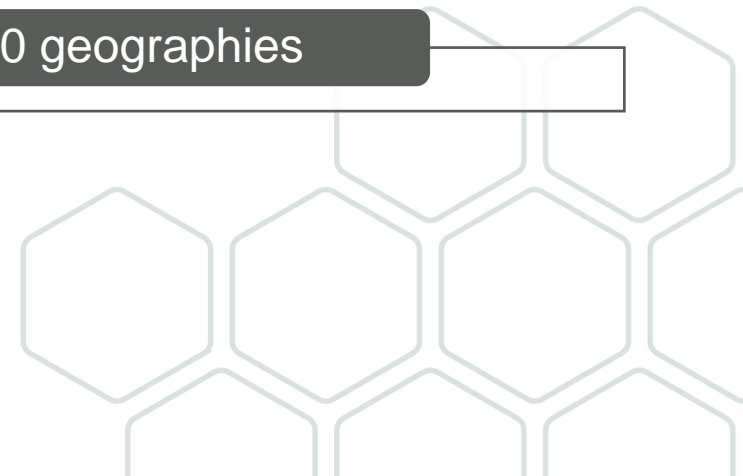


> 250GB data

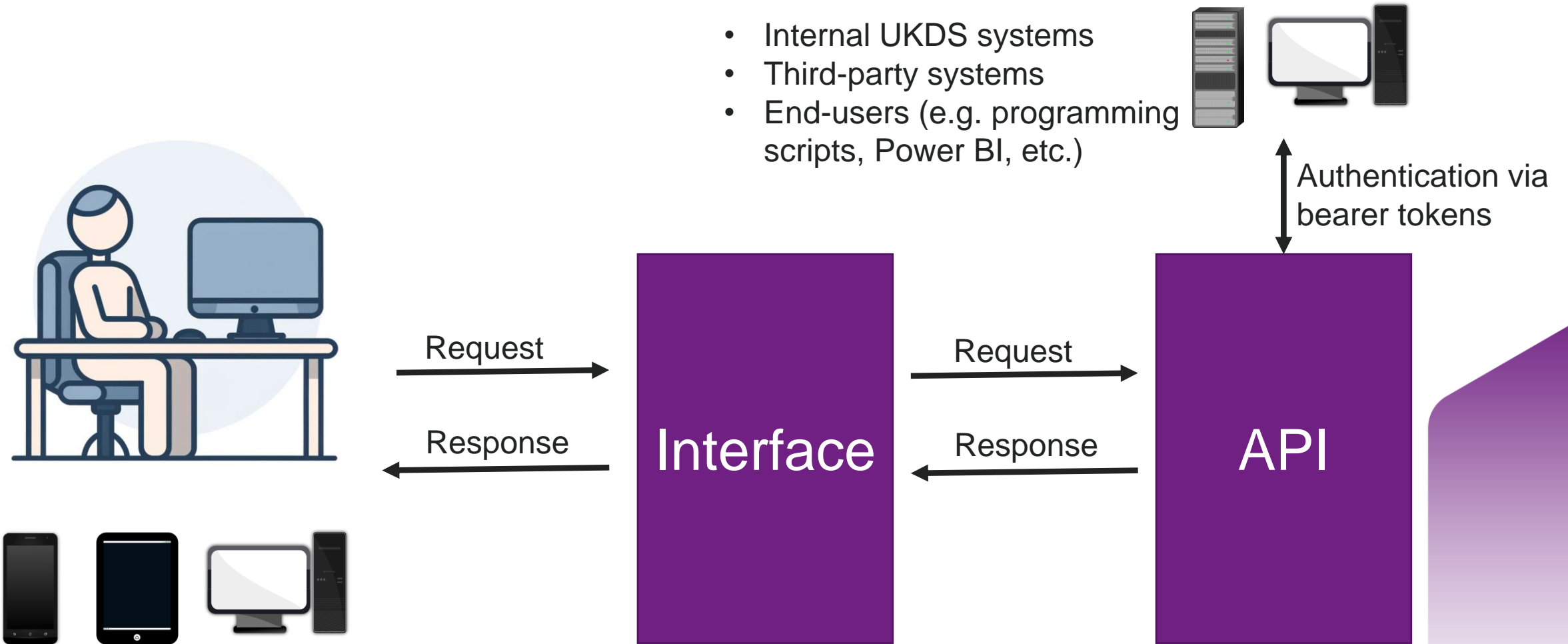
> 450 tables

> 250 topics

> 140 geographies



Platform access



REST API Specifications – OpenAPI 3

Representational State Transfer (REST) APIs are the most common type of APIs

It was preferred over other protocols like RPC, SOAP, and GraphQL.

The flow data API follows the OpenAPI 3 specifications.



Wide Adoption

Extensive
Support

Stateless
operations

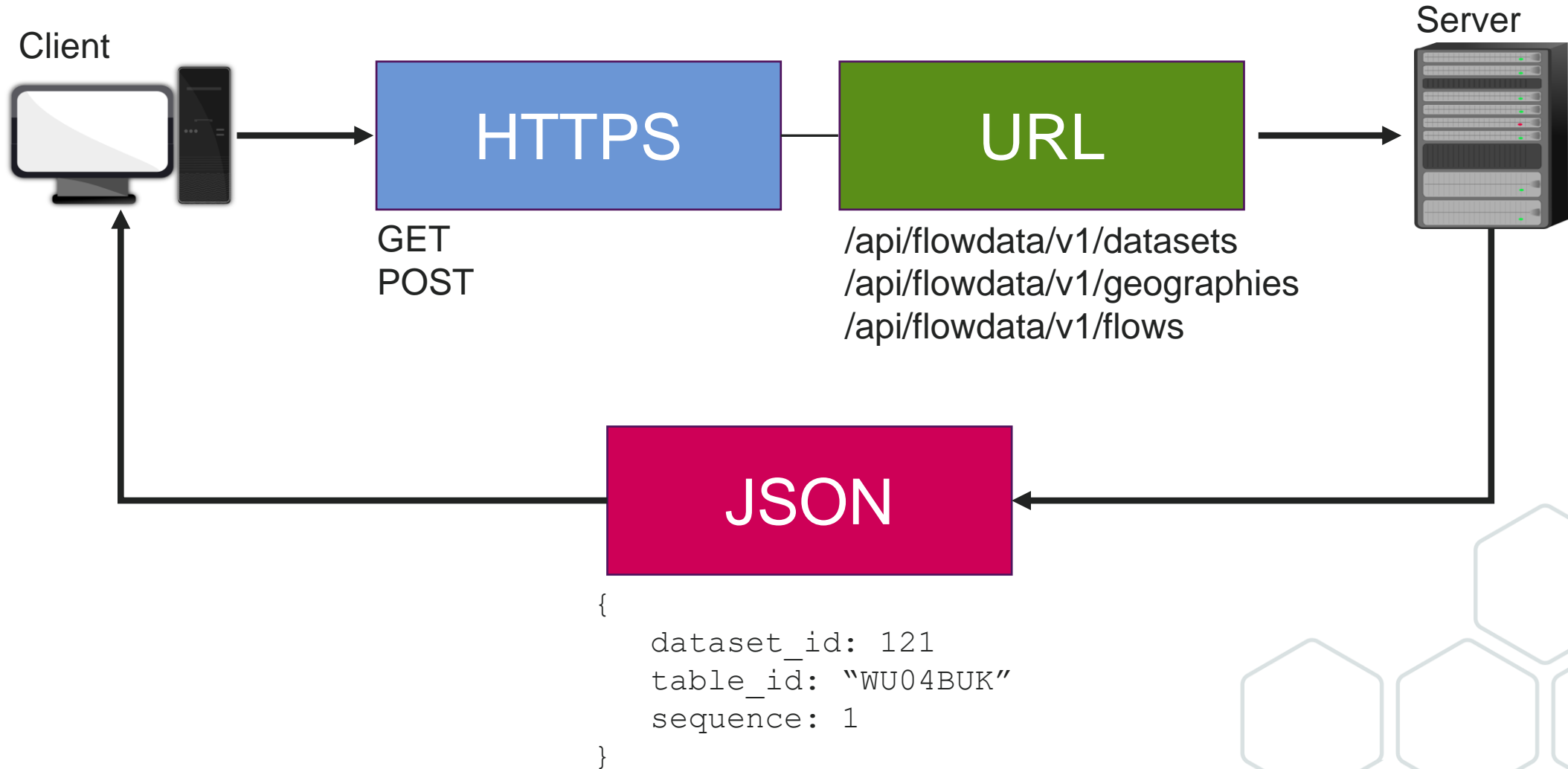
Interoperability

Easy
integration

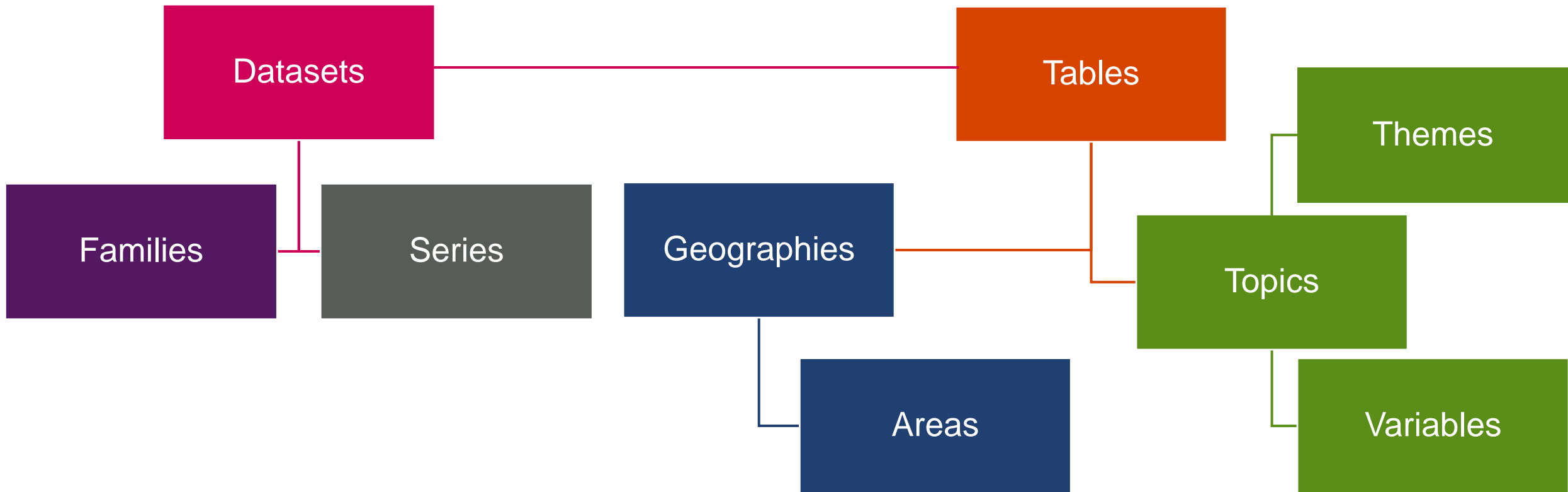
HTTP(S)
technology



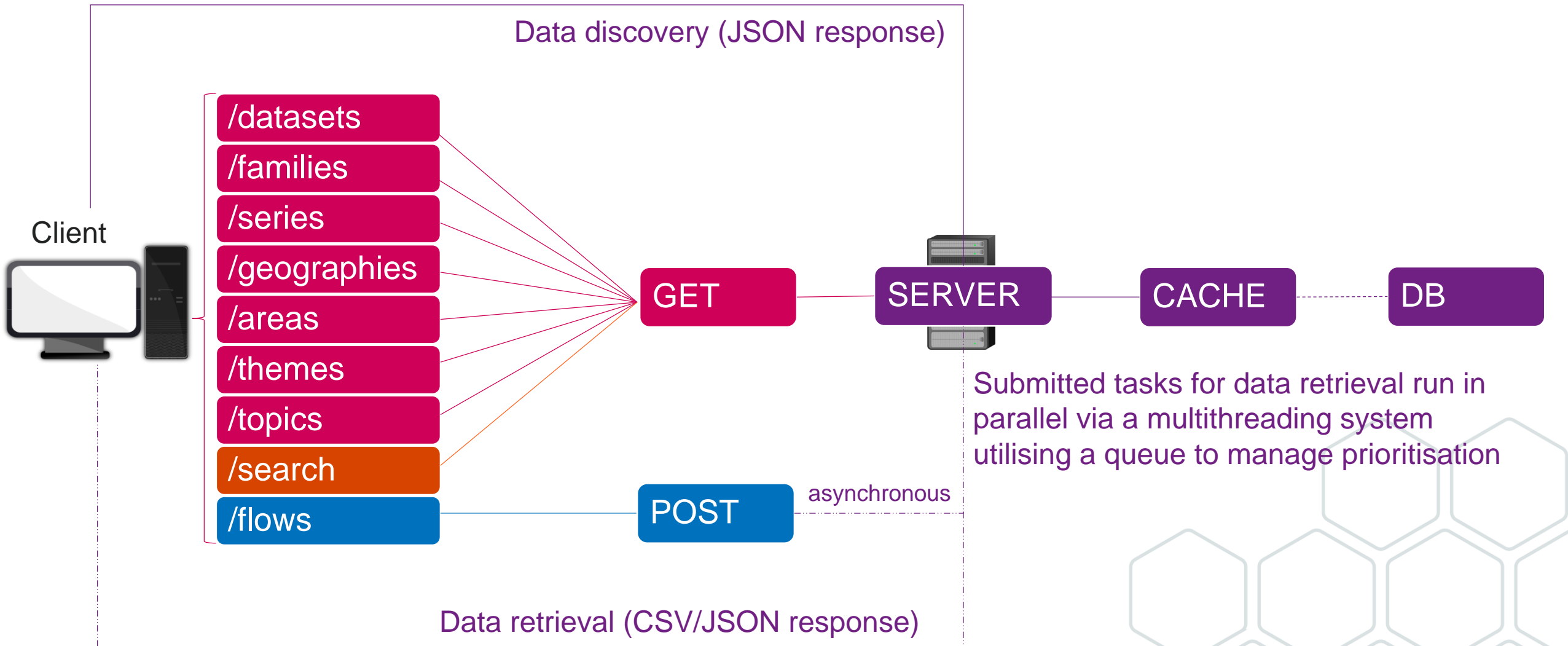
REST API



Census flow data taxonomy



Census flow data API endpoints - example



Example of available query parameters for /datasets

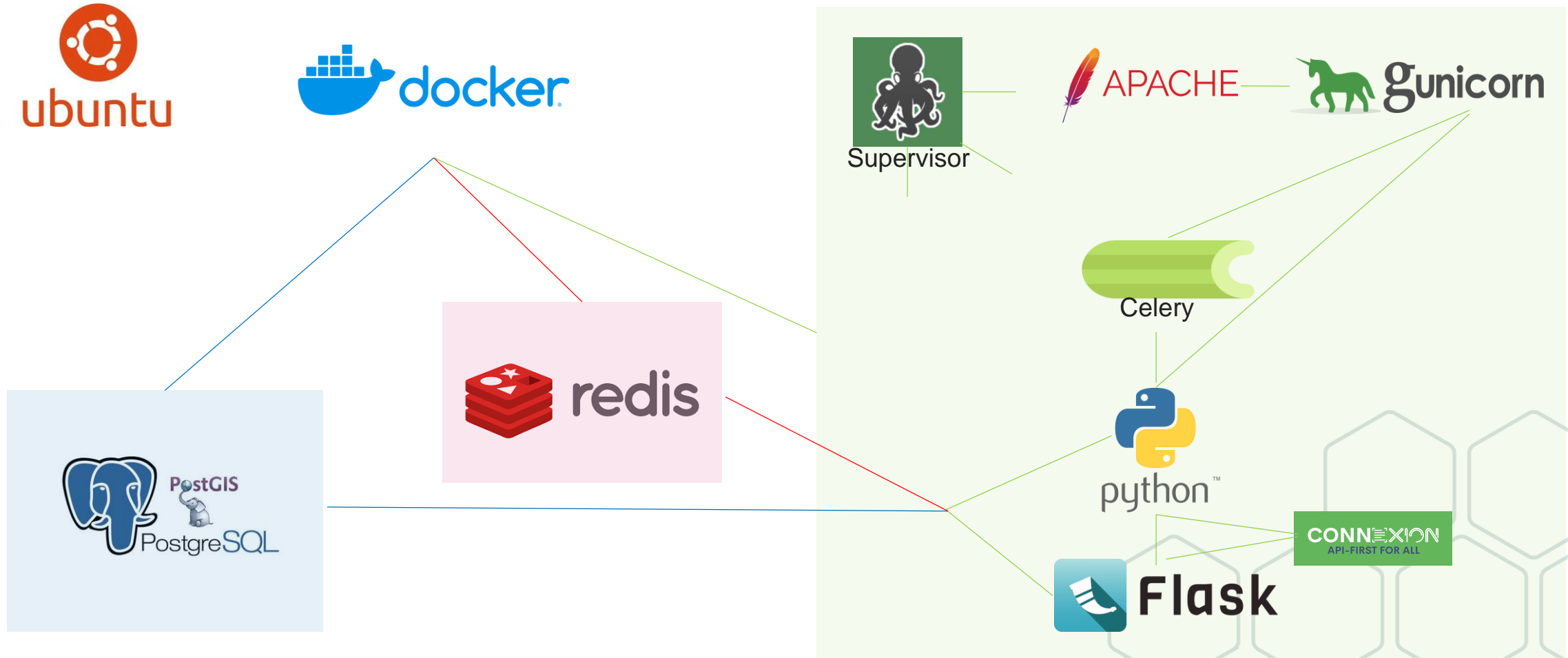
QUERY PARAMETERS

page	integer Example: <code>page=1</code> Number of page
results_per_page	integer Example: <code>results_per_page=20</code> Results per page
sort	Array of strings [1..3] items Example: <code>sort=geography_id:desc, table_id:asc</code> Sort the datasets by the given keys
group	Array of strings [1..2] items Example: <code>group=dataset_id, table_id</code> Group the datasets by the given keys
operator	string Example: <code>operator=any(geography_id,topic_alias);all(theme_id,family_id)</code> Defines the operator between parameters
match_type	string Example: <code>match_type=exact(geography_id,topic_alias);all(theme_id,family_id)</code> Defines the type of parameters (any, all, exact)
comparison	Array of strings [1..10] items Example: <code>comparison=dataset_id:gt, number_of_variables:bt</code> Allows to define a comparison operator for parameters (gt for greater than, lt for less than, eq for equal, bt for between). For bt there must be two parameters for the same identity, e.g. <code>dataset_id=10, dataset_id=20</code> .
fields_include	Array of strings [1..30] items Example: <code>fields_include=dataset_id, dataset_label</code> Allows to define the desired fields to be shown.
fields_omit	Array of strings [1..30] items Example: <code>fields_omit=dataset_id, dataset_label</code> Allows to define the fields excluded from being shown.
fields_add	Array of strings [1..30] items Example: <code>fields_add=geographies_full</code> Allows to show fields that are not shown by default
dataset_id	Array of integers [1..100] items Example: <code>dataset_id=10,11,12</code> Retrieve data by supported dataset id
geography_label	string Example: <code>geography_label=LSOA</code> Retrieve data by geography label
geography_id	Array of integers [1..10] items Example: <code>geography_id=10,11,12</code> Retrieve data by supported geography id
origin_geography_id	Array of integers [1..10] items Example: <code>origin_geography_id=10,11,12</code> Retrieve data by supported origin geography id
destination_geography_id	Array of integers [1..10] items Example: <code>destination_geography_id=10,11,12</code> Retrieve data by supported destination geography id
intrinsic_geography_id	Array of integers [1..10] items Example: <code>intrinsic_geography_id=10,11,12</code> Retrieve data by supported intrinsic geography id

aggregate_geography_id	Array of integers [1..10] items Example: <code>aggregate_geography_id=10,11,12</code> Retrieve data by supported aggregate geography id
topic_alias	Array of strings [1..10] items Example: <code>topic_alias=age1, carorvan1</code> Retrieve data by supported topic alias
topic_label	string Example: <code>topic_label=age</code> Retrieve data by topic label
row_topic_alias	Array of strings [1..10] items Example: <code>row_topic_alias=age1, carorvan1</code> Retrieve data by supported row topic alias
column_topic_alias	Array of strings [1..10] items Example: <code>column_topic_alias=age1, carorvan1</code> Retrieve data by supported column topic alias
theme_alias	Array of strings [1..10] items Example: <code>theme_alias=age</code> Retrieve data by theme alias
theme_label	string Example: <code>theme_label=age</code> Retrieve data by theme label
family_id	Array of integers [1..10] items Example: <code>family_id=1,2,3</code> Retrieve data by supported family id
series_id	Array of integers [1..10] items Example: <code>series_id=1,2,3</code> Retrieve data by supported series id
table_id	Array of strings [1..100] items Example: <code>table_id=MU03EW</code> Retrieve data by supported table id
table_title	string Example: <code>table_title=Resident</code> Retrieve data by table title
dataset_label	string Example: <code>dataset_label=Resident</code> Retrieve datasets by dataset label
dataset_description	string Example: <code>dataset_description=Resident</code> Retrieve datasets by dataset description

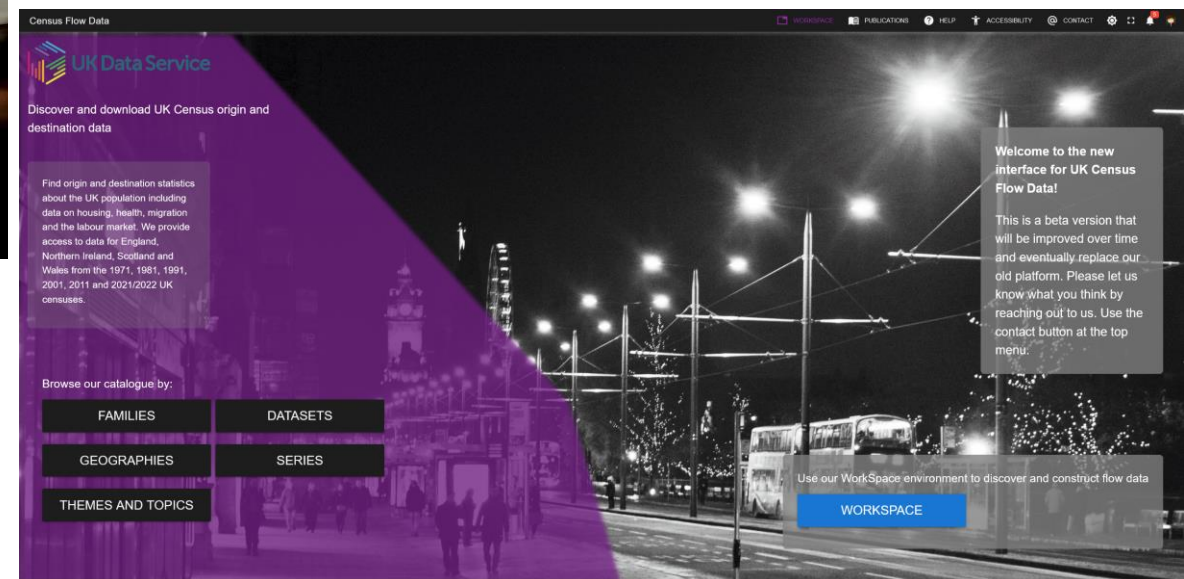
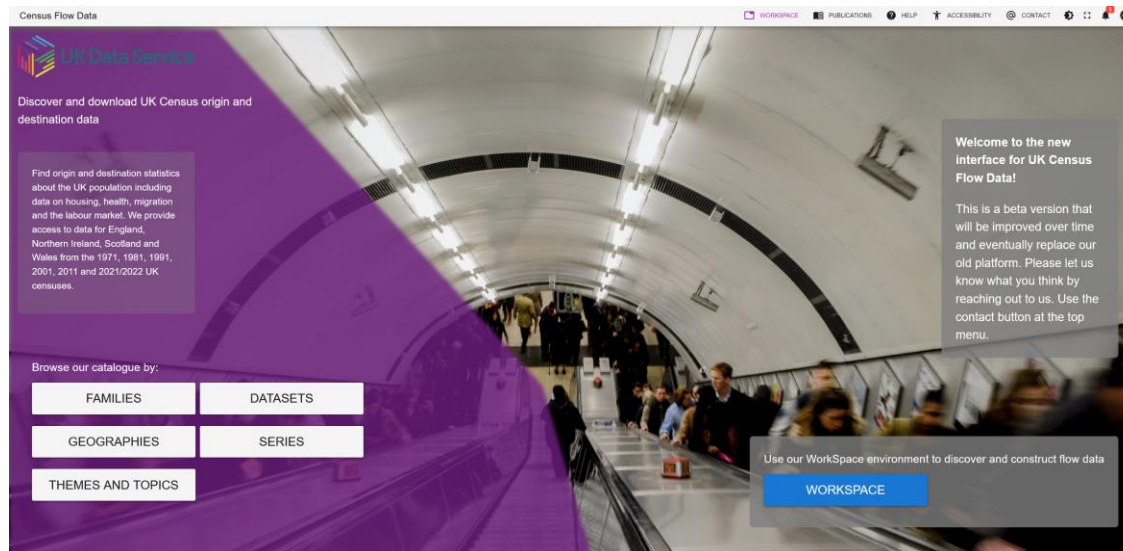


Technology stack of the API



Interface

We strive to make the new interface as modern and accessible as possible.



Features and enhancements

Not all functionalities will be released in the initial version of the platform. Some features, like areas selection via an interactive map, will be implemented gradually.

The API will offer new capabilities that were not previously available. For example, combining multiple tables on common geographical areas.

More work is needed to harmonise the metadata for all UK censuses from 1981 to 2021/22.



Lessons learnt



Building something from the ground up takes time. Prepare not to meet the original deadlines. Or the revised ones!



If you can apply for funding, do so! Having a funded project and a dedicated and experienced team to work on it will help immensely.



If something can go wrong, it probably will. Be ready to tackle unexpected challenges.



Thank you.

Vassilis Routsis, v.routsis@ucl.ac.uk

Oliver Duke-Williams, o.duke-williams@ucl.ac.uk

