



Wikibase for Smart Cities

A Roadmap

Houcemeddine Turki, Mohamed Ali Hadj Taieb, Mohamed Ben Aouicha



جامعة صفاقس
University of Sfax



WIKIMEDIA
TN User Group

Public university located in Sfax, Tunisia

- Top-ranked computer science research organization:
 - CTWS Leiden Ranking: 204th Worldwide, 1st in Africa
 - URAP Ranking: 482nd Worldwide, 2nd in Africa
- Top-ranked national university
- Sfax is the second largest town in Tunisia and is known for its nationwide leadership in public administration, industry and science
- Wikimedia-related Research has begun since 2012.
- We collaborate with user groups and affiliates mainly Wikimedia TN User Group, Wikimedia and Libraries User Group and Wikimedia Medicine in developing research projects of common interest.



About Us

Brief insights about the University of Sfax



Next steps

Future directions for our research group

We manage to create a new research unit

- The Research Unit is called *Data Engineering and Semantics*
- It will involve local Wikimedia researchers, Social media researchers and data scientists

Strengthening research efforts

- We are already involved in national and international research projects (DAAD)
- We try to tackle new projects



Introduction

The need to structured data in smart cities

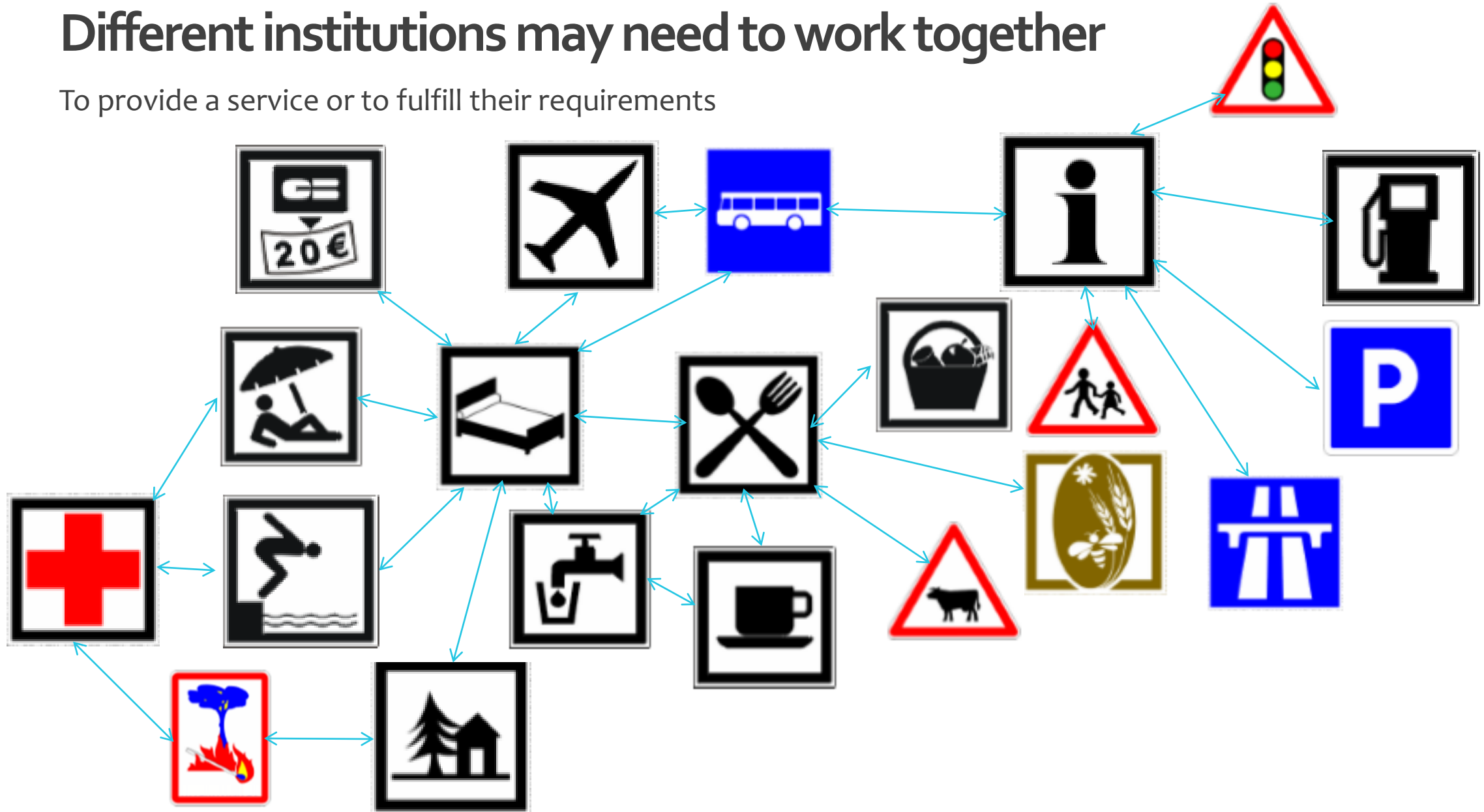
In a modern city, there is a huge number of institution types

Education, Industry, Health, Trade...



Different institutions may need to work together

To provide a service or to fulfill their requirements



However,

Many problems

- They use different data management software
 - Lack of interoperability
 - Lack of automation of the adjustment, alignment and update of databases
 - Paid service that requires regular updates and improvements
- Enterprise Data is not fully structured
 - Difficulty of the generation of statistical information from databases
 - Difficulty of quality assessment for enterprises
 - Difficulty of machine and deep learning of enterprise management and professional knowledge

Comparison

Current databases for enterprises vs. Structured data

Current data of enterprises

- Semi-Structured
 - Relational databases (RDBMS)
 - Data cannot be automatically known by machines
 - Queried using SQL, does not support federated queries
 - Data evaluation is difficult and requires knowledge in computing and the field of the concerned enterprise
 - Wikis (Mediawiki)
 - Automatic information retrieval and enrichment cannot be done from scratch by machines
 - Data validation and development is hard and requires an advanced knowledge of the architecture of the wiki, of wiki markup and of the Mediawiki API

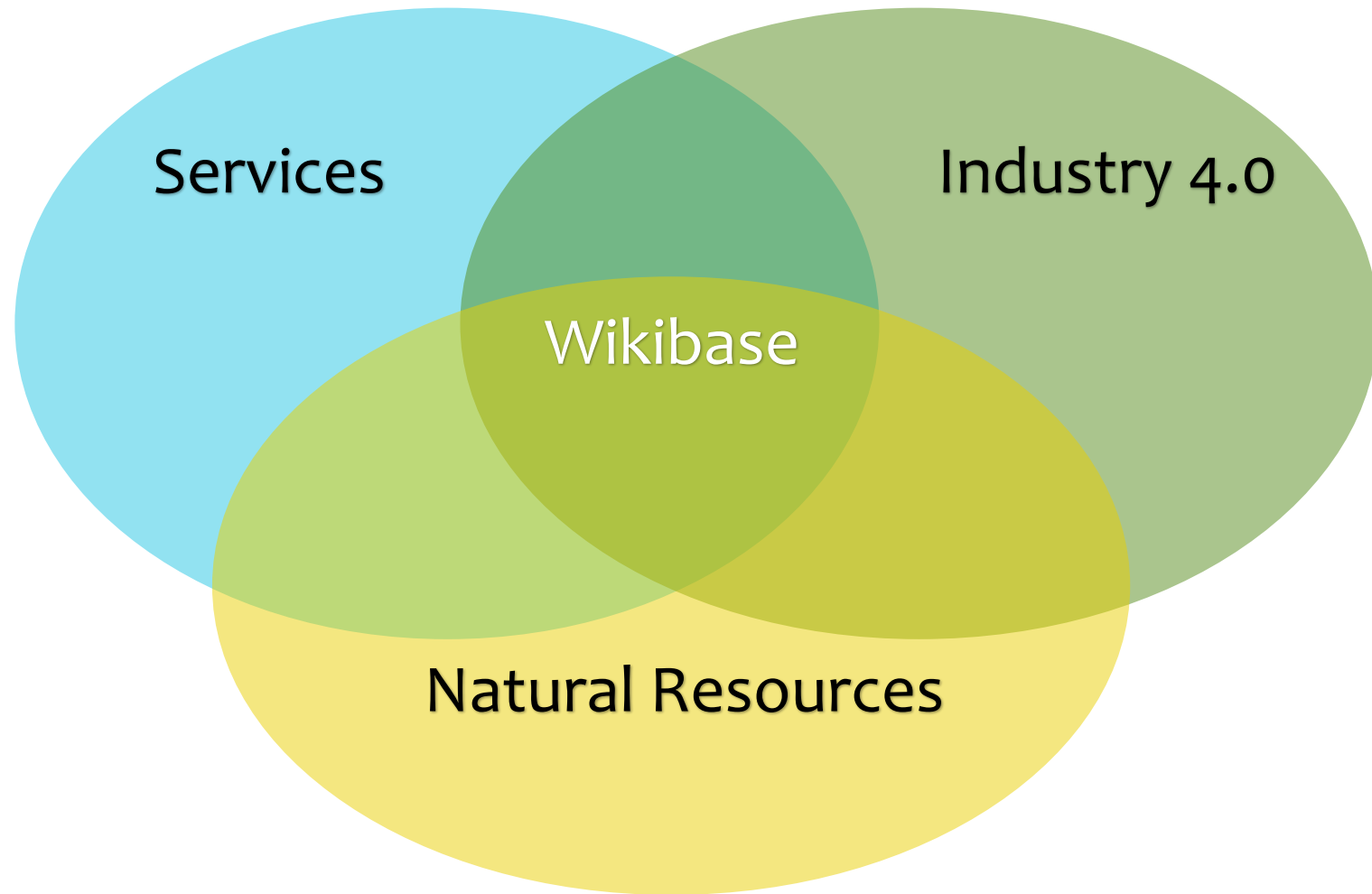
Structured data

- Fully structured in the form of triples (RDF)
 - Data can be recognized thanks to Description Logics
 - Queried using SPARQL, supports federated queries from various databases in RDF format
 - Data evaluation can be easily done by automatically integrating fully structured enterprise databases with a knowledge graph

Sfax, Tunisia as a Smart City

Current Status

- University of Sfax has launched a project to convert Sfax into a smart city
- Such a project needs a flexible data management software for multiple smart city purposes
- Wikibase as a free software can be an efficient solution particularly as Tunisia is still a developing nation



Wikibase

Definition and Features

WIKIBASE

What is Wikibase

From Wikibase Official Website

- Wikibase is an open-source software suite for creating fully structured knowledge databases, opening the door to the Linked Open Data web.
- Wikibase is the powerful, flexible and customizable knowledge base software that drives Wikidata. With MediaWiki as its front end and a rich JavaScript interface for data management, Wikibase makes collaboration easy for humans and machines alike, and its data model prioritizes language independence and knowledge diversity.
- Wikibase's robust data export options include JSON, RDF/XML, N3, YAML and more—and it provides a powerful query interface using SPARQL for both local and federated queries (with updates on changes), extending the potential data available from a single node to thousands.

Data Model

As seen from Wikidata, Wikibase-driven Knowledge Graph

- Entities are assigned a number beginning with a letter. The entities can be concepts or properties.
- Concepts can be:
 - Customers
 - Products
 - Employees
 - Departments
 - Sales
- Properties can be:
 - A type of relations between concepts
 - A type of relations between properties
 - A type of information about concepts or properties (e.g. Identifiers in other databases and Date of Birth)
- Each entity is assigned labels and descriptions in multiple languages
- Entities are linked to each others by relations in the form of triples that can include qualifiers and references.

Behcet's disease (Q911427)

rare immune-mediated small-vessel systemic vasculitis in humans

Adamanliades-Behcet disease | Behcet syndrome | Behcet's syndrome (disorder) | triple syndrome complex | Silk Road disease | Marbus Behcet | Morbus Behcet | Behcet syndrome | Adamanliades-Behcet disease | Behcet disease | Behcet's syndrome | Behcet's syndrome | Behcet's disease, Behcet's syndrome | Morbus Behcet's Syndrome | Behcet Disease | Behcet-Adamanliades Syndrome | Behcet syndrome | Behcet's Syndrome | Behcet's disease

Language	Label	Description	Also known as
English	Behcet's disease	rare immune-mediated small-vessel systemic vasculitis in humans	Adamanliades-Behcet disease Behcet syndrome Silk Road disease
Arabic	مرض بيهجت		متلازمة بيهجت داء بيهجت متلازمة بيهجت

Statements

- instance of** (disease) [1 reference]
- Designated intractable/rare diseases** [0 references]
- has cause** (immune system) [0 references]

Identifiers

- Orphan ID** (109950) [exact match]
- Disease Ontology ID** (DOID:13241)

Wikipedia (27 entries)

- ar:مرض بيهجت
- ca:Malaltia de Behçet
- de:Morbus Behçet
- en:Behçet's disease
- zh:貝賽特氏症

Wikibooks (0 entries)

Wikiversity (0 entries)

Wiktionary (0 entries)

Other sites (1 entry)

- commons:Category:Behçet's disease

User-Friendly Query Service

SPARQL Endpoint: <https://query.wikidata.org>

The screenshot shows the Wikidata Query Service interface. At the top, the browser address bar displays <https://query.wikidata.org/>. Below the browser, there is a search bar with the text "Enter text to search" and a "No results" button. The main interface features a "Query Helper" sidebar on the left with options like "Filter" (set to "main subject" and "2019-20 COVID-19 pandemic"), "Show", and "Limit 100". The central area contains a SPARQL query:

```
1 SELECT ?2019_20_COVID_19_pandemic ?2019_20_COVID_19_pandemicLabel WHERE {
2   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
3   ?2019_20_COVID_19_pandemic wdt:P921 wd:Q81068910.
4 }
5 LIMIT 100
```

Below the query, the "Result preview" section shows two columns: "2019_20_COVID_19_pandemic" and "2019_20_COVID_19_pandemicLabel". The results are as follows:

2019_20_COVID_19_pandemic	2019_20_COVID_19_pandemicLabel
Q83388131	The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China
Q83500389	Coronaviruses: genome structure, replication, and pathogenesis

Tools for automatic enrichment and validation of databases

API, QuickStatements, Bots, Wikidata Hub...

The screenshot shows the QuickStatements web interface. At the top, there is a search bar with the text "Find on page" and "Enter text to search". Below the search bar, there are several buttons: "New batch", "Last batches", "Chat", "Git", and "Help". A "Log in" button is also visible. The main content area contains a blue notification box that says "This is a new interface for QuickStatements V2. Just in case, the old interface is [here](#)." Below this, there are several input fields and buttons: "batch number" with a "See batch details" button, "user name" with a "See batches by user" button, and "temp. batch ID" with a "Discuss/revert a temporary (browser-based) batch" button.

The screenshot shows the MediaWiki API documentation page. The URL is "https://www.wikidata.org/w/api.php". The page content includes a search bar, a description of the page as an "auto-generated MediaWiki API documentation page", and a list of links: "[Documentation · FAQ · Mailing list · API Announcements · Bugs & requests]". The "Status" section states: "The MediaWiki API is a mature and stable interface that is actively supported and improved. While we try to avoid it, we may occasionally need to make breaking changes; subscribe to the [mediawiki-api-announce mailing list](#) for notice of updates." The "Erroneous requests" section states: "When erroneous requests are sent to the API, an HTTP header will be sent with the key 'MediaWiki-API-Error' and then both the value of the header and the error code sent back will be set to the same value. For more information see [API: Errors and warnings](#)." The "Testing" section states: "For ease of testing API requests, see [Special:ApiSandbox](#)." The "Parameters" section lists several parameters: "action" (Which action to perform.), "abusefiltercheckmatch" (Check to see if an AbuseFilter matches a set of variables, an edit, or a logged AbuseFilter event.), "abusefilterchecksyntax" (Check syntax of an AbuseFilter filter.), "abusefilterrevalexpression" (Evaluates an AbuseFilter expression.), "abusefilterunblockautopromote" (Unblocks a user from receiving autopromotions due to an abusefilter consequence.), and "abuselogprivatedetails" (View private details of an AbuseLog entry.).

Files describing the expected features for a given type of entities

ShEx

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

start = @<human>

<human> EXTRA wdt:P31 {
  wdt:P31 [wd:Q5];
  wdt:P21 [wd:Q6581097 wd:Q6581072 wd:Q1097630 wd:Q1052281 wd:Q2449503 wd:Q48270]?; # gender
  wdt:P19 . ?; # place of birth
  wdt:P569 . ? ; # date of birth
  wdt:P735 . * ; # given name
  wdt:P734 . * ; # family name
  wdt:P106 . * ; # occupation
  wdt:P27 @<country> *; # country of citizenship
  wdt:P22 @<human> *; # father
  wdt:P25 @<human> *; # mother
  wdt:P3373 @<human> *; # sibling
  wdt:P26 @<human> *; # husband/wife
  wdt:P40 @<human> *; # children
  wdt:P1038 @<human> *; # relatives
  wdt:P103 @<language> *;
  wdt:P1412 @<language> *;
  wdt:P6886 @<language> *;
  rdfs:label rdf:langString+;
}

<country> EXTRA wdt:P31 {
  wdt:P31 [wd:Q6256 wd:Q3024240 wd:Q3624078] +;
}

<language> EXTRA wdt:P31 {
  wdt:P31 [wd:Q34770 wd:Q1288568] +;
}
```

Easy install

<https://www.wikiba.se/install/>



Showcase

Install

Resources

FAQ

Contact

Quick start with Docker

Introduction

We've put together a set of machines in Docker that should have you up and running in no time. This configuration starts an empty instance of Wikibase, a MediaWiki front end with query interface, a [query backend](#), Elasticsearch, and a QuickStatements bulk editing service.

Before you start

- You'll need to have [docker](#) and [docker-compose](#) installed on the computer where you want to run your Wikibase instance.
- Together, these Docker machines require at least 4GB of memory.
 - Check out Docker's [documentation on resource constraints](#).
- An empty Wikibase running on Docker requires at minimum 6GB of disk storage.

Getting the machine images running

1. Download the [docker-compose file](#) and place it on the computer where Docker Engine and docker-compose are installed.
2. In the directory that now contains the `docker-compose.yml` file, run the following to pull the needed Docker machine images:

```
docker-compose pull
```

3. Start the machine images in the background:

```
docker-compose up -d
```

(To view the continuous output, you can run `docker-compose logs -f`)

4. Verify that all the services have started.

1. Run `docker-compose ps`. You should see eight images in an "Up" state.
2. Check the logs for success or errors -- for example, using this command:

```
docker-compose logs --tail="20" -t
```

3. Try to load the front end and the query interface in your browser (see below).

To sum up

- Easily readable by machine, can be used in updating other databases
- Easily verifiable and editable by machine
 - Can be simply enriched using Knowledge Graph Learning (from databases, raw texts, IoT...)
 - Can be simply validated using federated queries and description logics (ShEx, statements related to properties...)
- Free and available online at <https://wikiba.se/>.
 - Does not require subscription
 - Updated by many volunteers
 - Easy install



Challenges

How to define Wikibase applications for smart cities

Matters

- Not all needed codes are available at <https://wikiba.se/>.
 - Useful extensions that are developed to enhance Wikidata can be useful for other Wikibase-driven databases.
 - However, they should be gathered from many GitHub repositories.
- When Wikibase-driven databases are many in a city, alignment between them is absolutely difficult
- Several types of businesses need specific data models
 - A databases for sales does not have the same structure as a database of health services
- Adequacy of the use of big data technologies (Hadoop, MapReduce and Spark) for the enrichment and refinement of Wikibase applications is not well documented
- Security and upgrade of Wikibase and Mediawiki software for enterprise applications is perplexing
- Conflict of interest with Wikidata
 - Several data can be relevant for inclusion in Wikidata: Data of significance for a broad community such as a database of cultural heritage
 - Several data can be useless for Wikidata such as the transactions of an individual. However, their synthesis can be useful for Wikidata

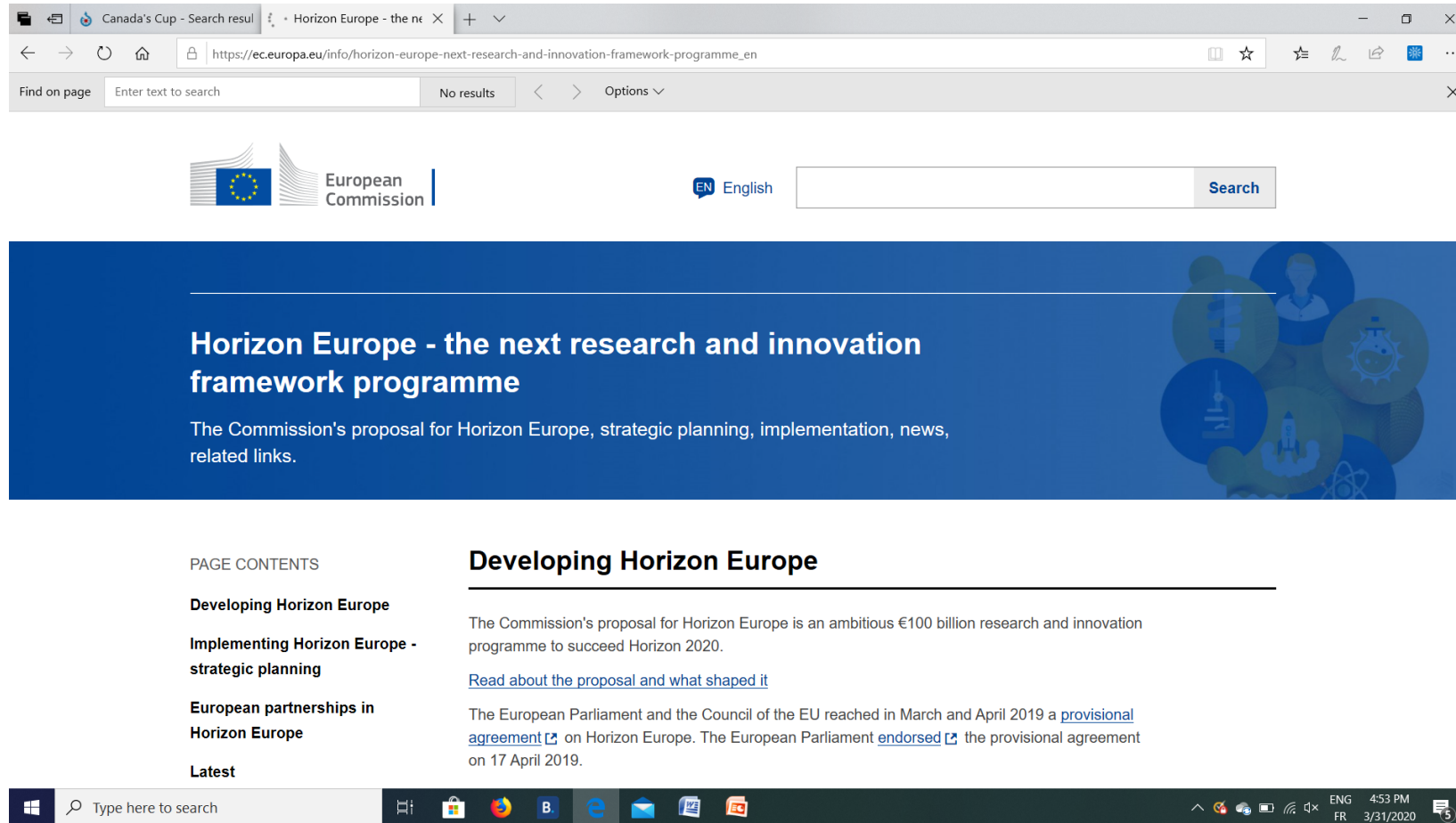
Solutions

- Let all useful tools for a Wikibase-driven application in a unique repository
- Define data models for each type of businesses: Pharmacy, Grocery, Market, Industry...
- Define when data can be directly uploaded to Wikidata
- Tools should be developed so that Wikibase applications for enterprises can benefit from useful information available in big data (IoT, Social Media) and Wikidata
- Tools to upgrade Wikibase and Mediawiki software and to ensure security of smart cities applications should be developed
- Create a skeleton of a Wikibase-driven database for the alignment between Wikibase applications in a smart city

However, funding is required...

We manage to participate to Horizon Europe in 2021.

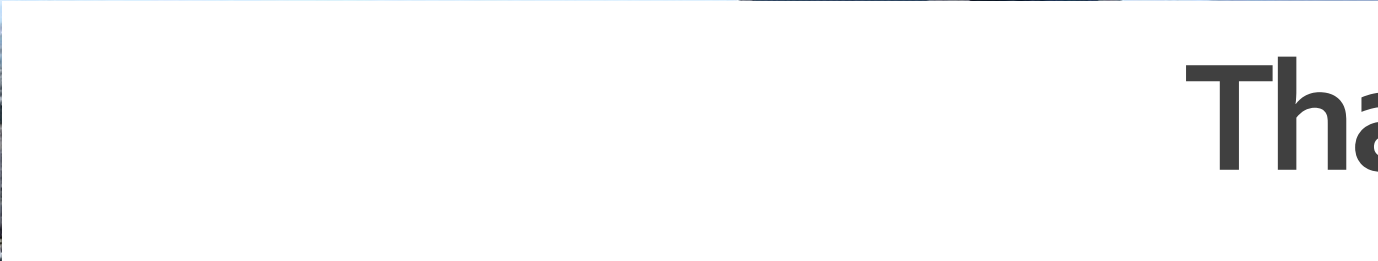
All corporations and volunteers who would like to join our open society project are greeted.



The screenshot shows a web browser window displaying the European Commission's page for Horizon Europe. The browser's address bar shows the URL: https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en. The page features the European Commission logo and a search bar. The main heading is "Horizon Europe - the next research and innovation framework programme". Below this, a sub-heading reads: "The Commission's proposal for Horizon Europe, strategic planning, implementation, news, related links." The page is divided into two columns. The left column, titled "PAGE CONTENTS", lists: "Developing Horizon Europe", "Implementing Horizon Europe - strategic planning", "European partnerships in Horizon Europe", and "Latest". The right column, titled "Developing Horizon Europe", contains the following text: "The Commission's proposal for Horizon Europe is an ambitious €100 billion research and innovation programme to succeed Horizon 2020." followed by a link: "[Read about the proposal and what shaped it](#)". Below this, it states: "The European Parliament and the Council of the EU reached in March and April 2019 a [provisional agreement](#) on Horizon Europe. The European Parliament [endorsed](#) the provisional agreement on 17 April 2019." The Windows taskbar at the bottom shows the date as 3/31/2020 and the time as 4:53 PM.

References

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 34-43.
- Wikibase (2020). *Wikibase Official Website*. Accessible at <https://www.wikiba.se/>.
- Turki, H., Shafee, T., Taieb, M. A. H., Aouicha, M. B., Vrandečić, D., Das, D., & Hamdi, H. (2019). Wikidata: A large-scale collaborative ontological medical database. *Journal of Biomedical Informatics*, 99, 103292.
- Thornton, K., Solbrig, H., Stupp, G. S., Gayo, J. E. L., Mietchen, D., Prud'Hommeaux, E., & Waagmeester, A. (2019, June). Using shape expressions (ShEx) to share RDF data models and to guide curation with rigorous validation. In *European Semantic Web Conference* (pp. 606-620). Springer, Cham.



Thank you

turkiabdelwaheb@hotmail.fr
mohamedali.hadjtaieb@gmail.com
mohamed.benaouicha@fss.usf.tn

+21629499418

