SEADEV Presentation

RESOURCES, PROJECTS & SERVICES

GILBERT BRAULT

Contents

Context

Sea Dev was created in 2020 by Gilbert Brault ([LinkedIn](#)).

The purpose of Sea Dev is to develop applications with a significant business impact using the Jupyter development framework, transforming company data into business knowledge, implementing daily tools for business users. The core technology of applications is software.

In 2021, Sea Dev has developed multiple projects for the following companies

- Vink Chemicals GmbH
    - An audit and management framework (delivered)
    - A product selector application (delivered)
- Ajanla Farms Ltd
    - Poultry Flock management reports (in-development)

And has started the following R&D developments

- ARUCO based asset tracking management (in-development)
- Excel automation: Tabular data extraction and automated pivot table creation (delivered)
- Business Process: Organizing the Business Process Modelling Notation (BPMN) Graphs in collections (applications) and keep their textual documentation in sync, to become the single version of the truth, companywide.

Why should we work together?

Our motto is "Improve your business knowledge base on your existing data".

Your company has existing data sources which are seldom fully used in your business processes: we can help you identify the greatest potential of those sources for your business.

We then deploy the appropriate technology to transform those assets leveraging the efficiency of your company employees in their daily work as live information will raise their awareness and knowledge.

Bottom line, this will improve your operational margins!

Those actions are performed in short life cycle projects which yield results in weeks!

Resources and Knowledge

People

	Gilbert Brault - 30-year experience in R&D, Marketing and Strategy at Schneider Electric - PhD in Industrial Automation
 gbresume.pdf	Please, see attached document

Technology

The technology know-how of Sea Dev is based upon

- Data modeling and algorithm research impacting company business process (data driven company – Digital age)
- Widely reusing existing technology (Jupyter project, open-source software, cloud computing)
- Practical experience to design
  - Business process
  - Implement them using leading edge computer tools
- Computer technology
  - Communication: digital radio, network, IP technology, W3C technologies (HTTP…)
  - Data modeling (including SQL database)
  - Processing languages (Python, Java, C#, C…)
  - IoT (Internet of Things)
  - Virtualization and cloud computing
  - Optical Code Recognition (OCR)
  - Document Format transform (ex: HTML -> PDF, Text -> Word etc…)
- Designing specification, documentation, and training

Project briefs

Ilex Ascenseurs: Digitizing Elevator Control Reports

Ilex Ascenseurs is a medium-sized French company, headquartered in Antibes with an annual turnover between 15-20M€ and 150-200 employees in France. Ilex Ascenseurs specializes in elevators: new installations and maintenance (installed base of ~10,000 elevators). The total installed base in France amounts to around 500,000 lifts.

The problem to solve

The elevator business in France is fairly regulated as an equipment must be controlled every 5 years. Specialized Controlling Engineering firms are checking units are complying with the legal framework, released by the French government.

An audit report is released which is composed of 5 to 10 pages per unit. Ilex Ascenseurs is processing more than 2,000 reports a year (~20,000 to 40,000 report pages).

Among this report, only the "observations" are of interest for Ilex Ascenseurs. Those notices are requests made by the auditor to carry some work and actions to fix issues, to make the unit under test, compliant with regulation.

Each Auditing firm has its own report layout. Ilex Ascenseurs is currently dealing with more than 15 auditing firms.

The current process is the following:

- Receiving pdf files by email from the real estate companies
- Building manually Excel files, extracting the relevant information from the pdf (cut and paste)
- Using the excel file as a vehicle to monitor corrective actions progress and assign work to employees
- Managing windows OS repositories to keep track of the above information.

The amount of manual work and the dispersion of information makes this process error prone and increase the processing lead time. It is also man-power intensive, to provide the actual status for a given equipment fixes, to the real estate agent managing a given property.

The solution
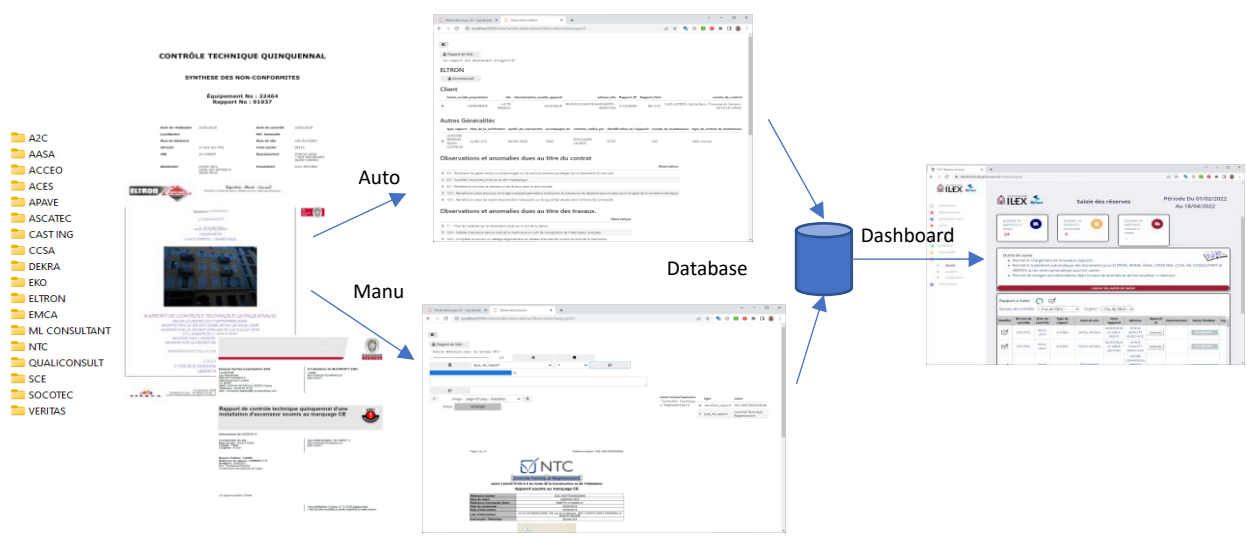
- Email received from the real-estate managing company (no-change)

1. A web application (the digitizing module), accessible to selected Ilex employees from the internet, is translating the audit bureau pdf report automatically or manually. The extracted information is stored in a database. The original file is stored for further retrieval.
2. The database is shared with a dashboard application (global to Ilex Ascenseurs), which, for the auditing reports, is organizing the workflow:
   a. Data entry (Dashboard application launch the digitizing module) -- Saisie
   b. Work organization and scheduling -- Analyse
   c. Progress monitoring -- Traitement
3. The automatic mode is available for the most frequent reports (to date 7 Auditing firms).
4. The manual mode is an assisted Optical Code Recognition (OCR) enabling construction of all the database items selected one after the other by the user (tagging).
5. The manual mode is used for any Auditing Firms which are not automated and for the reports which are not pdf native.

Benefits
1. Workflow Integration into Ilex dashboard (unified access for employees).
2. A unified database to hold all the received reports and transforming them into structured data.
3. Automated translation enabling fast and accurate translation of all observations without report specific knowledge.
4. Assisted OCR for manual cases
5. Extensive access to original reports information, corresponding scheduled actions, and monitoring thanks to a searchable database
6. Retrieval of the original reports for cross-checking purposes
7. Overall performance improvements
   a. lead-time (from days to hours)
   b. workload decrease (from 1h/report to 10 to 30 minutes Auto/Manu)
   c. information retrieval (from hours to minutes)

Technology
Some of the salient technical features are:

1. Jupyter Notebooks running into voila run-time
2. Tesseract 5.0 for OCR (French specialization)
3. Region Of Interest module form Biophysics @ TU Wien (to capture manual data)
4. MySQL database
5. PDF processing library (pymupdf) to read and chunk pdf documents
6. All specific software is written in the Python language

The text classification, which has been agreed with Ilex specialists, is performed by setting the appropriate section delimiters based upon either position and/or headers contents. Regular expressions patterns are sometime used for section delimiters recognition.

Report uniqueness (to check a report is processed only once) is implemented making a hash of the report file.

A global procedure is designed to have all the various reports analyzed according to a unique pattern. This pattern is captured as a generic python class. Each auditing report format is specialized implementing a derived class per audit firm or report format.

With this design, implementation of a new report extraction is now set to 2 to 3 days (maybe even as low as 1 day).

Vink Chemicals Audit & management framework

Vink Chemical (100 people, 50M€ turnover), based in Germany, is a fast-growing company, created in 2011, providing biocides to various markets segments including Paint & Coating, Oil & Gas, Personal Care...
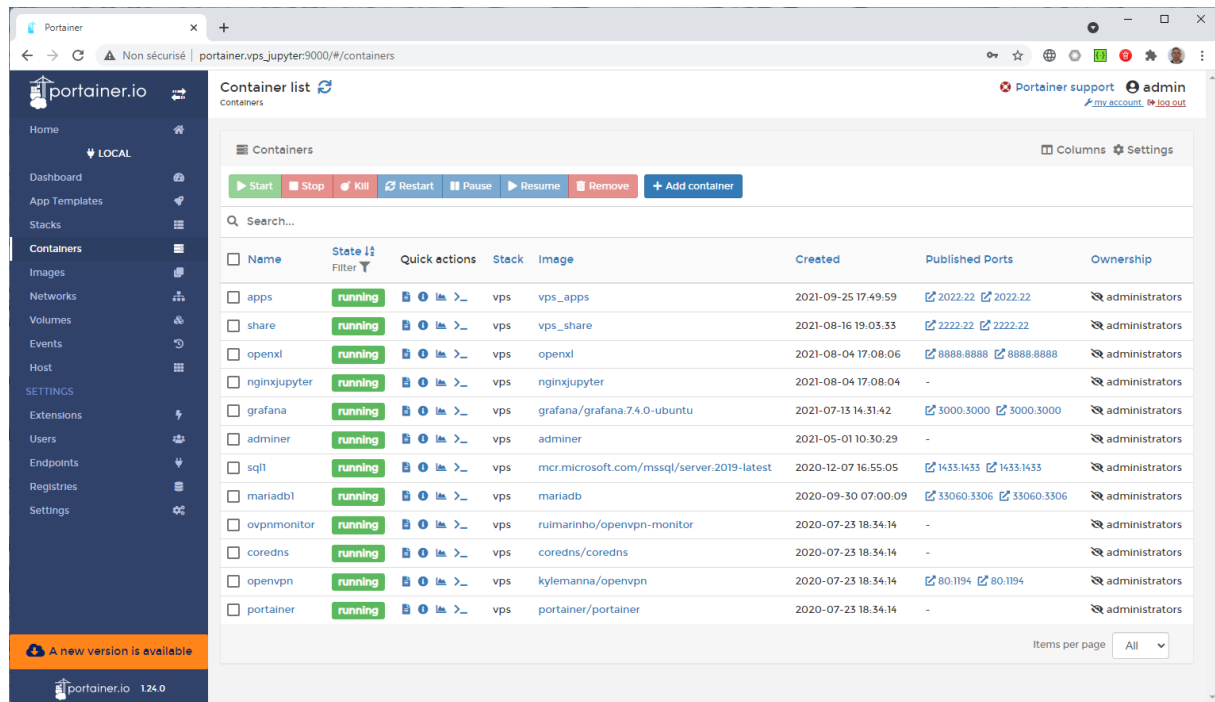

Figure 1: Portainer Docker Stack deployment and monitoring

The problem to solve

The company is using an ERP tool (Enterprise Resources Planning) and at the same time, line managers are using a lot of excel intermediate files.

 The issue is then the duplication of information which provides business problems as there are many versions of the truth.

The solution

1. Identify the usage of the most relevant excel files per business center
2. Makes the ERP data accessible
3. Design and implement a process running daily which
    a. Use the data from the ERP
    b. Creates excel files in line with the business practices
    c. Deliver them in a SharePoint to be easily accessible by the business users

Benefits

1. Using excel to infuse the company with ERP data to close the multiple versions of the truth gaps
2. Excel is an easy-to-use tool, compared to the ERP and more company employee have access to data guiding their daily work.
3. Special Processing functions take care of issues which were not supported by the ERP (e.g.: product cost calculation per batch)

4. As the company data is now 'open', the capability to generate on demand analysis is possible (for example sales forecasting based on historical invoicing).

Technology

1. A Docker enabled VPS (Virtual private server) including
   a. A Jupyter notebook server executing a daily processing (cron[1])
   b. A database server which replicates the ERP database
   c. A management server to monitor the virtual stack (Portainer)
2. A windows 2019 Server supporting
   a. The database import scripts as the ERP technology impose windows
   b. The excel pivot generation (as excel just runs on windows)
3. A SharePoint space collecting the generated excels
   a. Generated pivots are emailed to Office 365 SharePoint

This stack is accessible via a VPN, implementing the application protection layer and which guaranty that only selected users have access to the resource.
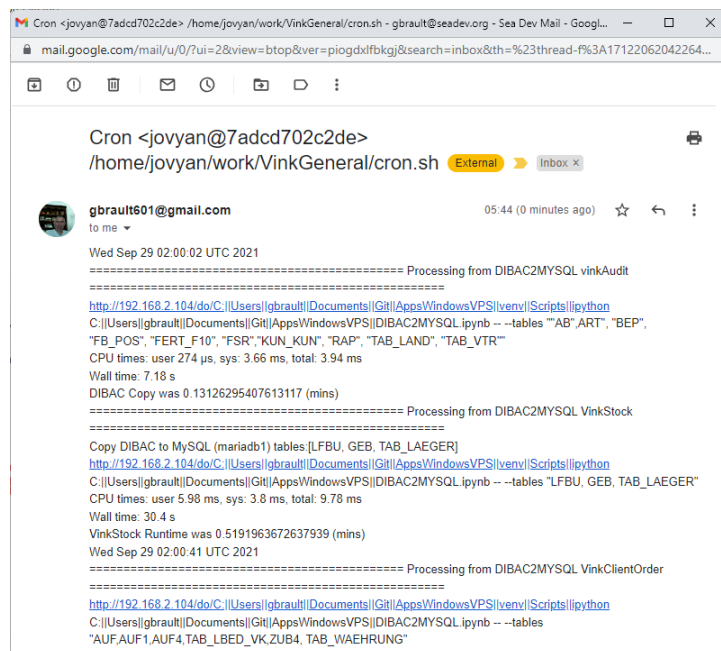

Figure 2: Daily cron execution report

---

[1] A cron is a Linux daemon which execute task according to a specified schedule

Vink Chemical Product Selector

Biocide's usages depend on multiple factors and the company must guide users, selecting the product corresponding to their requirements.
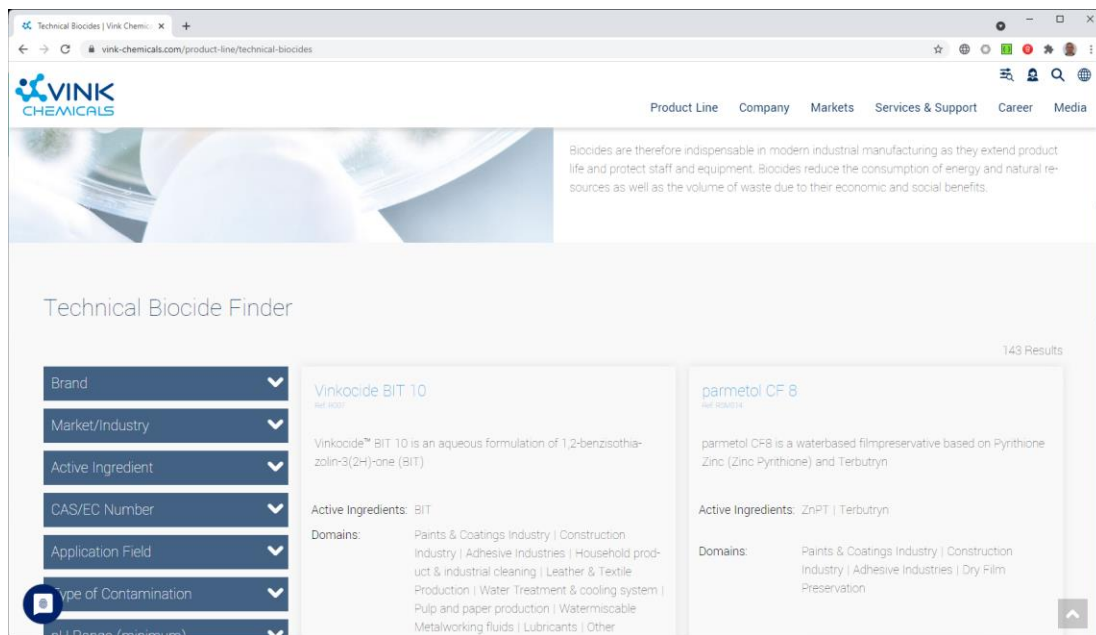

Figure 3: Parametric Product Selection

The problem to solve

Help sales and communication department accurately define the main descriptors for biocides, considering the stakeholders from customer organization and inside Vink chemicals.

Clean the product portfolio to propose an Up-To-Date catalog including Vink Chemical legacy and the new acquired line of products.

Get this data accessible from Vink Chemical corporate website as a first milestone.

The solution

1. Creation with Vink Chemical project owner of a unique excel file, defining all the catalog clustering data, and providing products main specifications, identification, and presentation
2. The excel file content is driving the data clustering and the User Interface content (clusters)
3. Define a standard interface (REST API) to ease the integration into the corporate website done by an IT partner.
4. Monitor the implementation regularly to get the goal achieved

Benefits

1. The sales and communication department are the owner of the data and the way it is distributed and presented thanks to this governing excel file which is automatically uploaded
2. An easy to search product catalog, based on the product definition content

Technology

1. The REST API is based on python Flask technology
2. "OpenAPI" browser, generated from the code definition to help IT partner integrate the product selector into the corporate website
3. A virtual stack, used to deploy and manage the application deployed on a VPS (Virtual Private Server)

Poultry Flock management reports

 Ajanla Farms Ltd, based in Nigeria, is breeding poultry for eggs or meat markets.
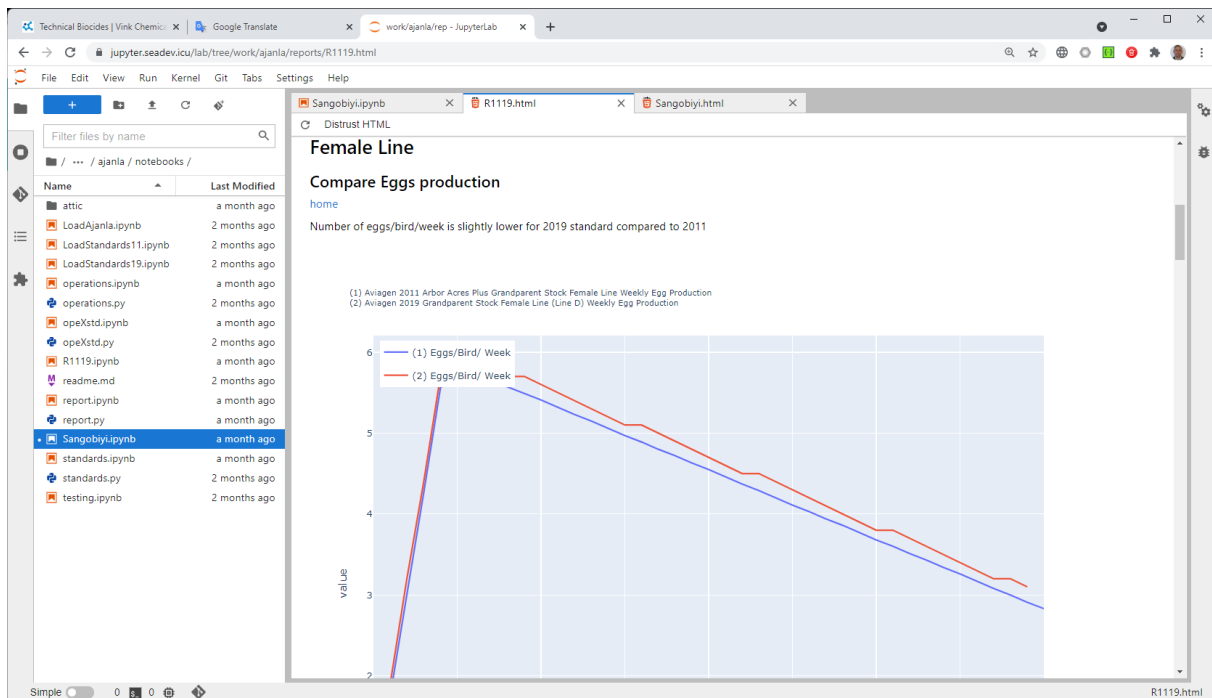

Figure 4: Compare Standards with Operation

The problem to solve

Management today receives excel files which are pure tables of numbers. It is cumbersome to read and difficult to find any insights from this raw data. The process of data collection and reports needs also to be improved.

The solution

Design a process to transform the current situation into a live data collection and processing practice.

1. The first step is to model the data to be able to compare performance with goals set by the upstream producers.
2. Then propose periodical reports which are designed to highlight issue or performance KPIs, using graphic representation for instant understanding of data
3. To achieve this level of report, a good data capture must be designed and agreed, plus a full defined business process must be agreed among the company stakeholders.

Benefits

The main expected benefit is to enable management taking decision based on factual data interpretation and in a timely manner.

Technology

The mockup is based on a VPS[2] stack, the user interface is based on Jupyter framework

---

[2] VPS: Virtual Private Server

ARUCO based asset tracking

In many industrial cases, asset automated identification and location is a problem.
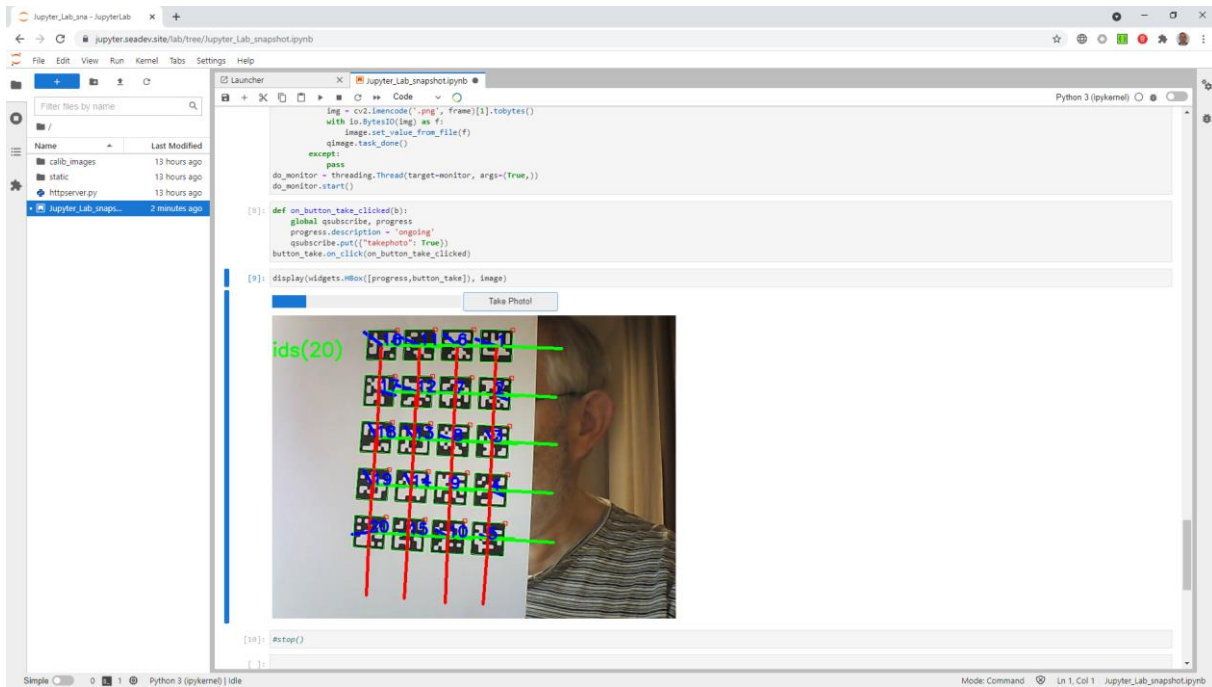

Figure 5: ARUCo, many markers can be detected simultaneously

The problem to solve

Two use cases are central for this development

- Maintenance team 'augmented reality' (production equipment tracking)
- Material asset tracking (production goods tracking)

Maintenance use case:

- A technician in field operation needs to quickly find the documentation about a piece of equipment
    - Having a camera that scans the scene, ARUCO markers are gathered
    - Transformed into messages
    - Messages are used to identify assets and get their position
    - Identifiers are used to search database to find documentation

Production use case

- Packing material is decorated with ARUCO signs
- A camera (or many cameras) scans the scene
- Position and identification of material is then carried
- According to the task, the derived information enables to track the production process or locate material or facilitate stock and storage management

The solution

We are at the specification and experimentation phase for this project, but we believe ARUCO technology can be a good solution in this field

ARUCO signs have been developed in the last decade and stands for "Augmented Reality University of Cordoba":

"Speeded up detection of squared fiducial markers", Francisco J.Romero-Ramirez, Rafael Muñoz-Salinas, Rafael Medina-Carnicer, Image and Vision Computing, vol 76, pages 38-47, year 2018

OpenCV library implements ARUCO processing.

Here is an ARUCO marker

	Depending upon the selected 'dictionary', markers can represent up to 1000 different 'Ideograms'. When searching for Ideograms from a given dictionary, other ideograms from other dictionaries are ignored. Recognition speed of ARUCO markers is very fast. ARUCO markers are used for drone missions.

The idea is to compose messages with ARUCO markers and develop a library based upon OpenCV to extend the use of ARUCO to identification.

Benefits
- ARUCO markers can be read in multiple positions
- ARUCO markers are detected in parallel (many ARUCO markers in the scene)
- ARUCO markers have redundant definition to avoid errors
- ARUCO markers can be used to find the position of assets
- A handful of identifiers can be scanned with relatively cheap equipment covering a large scene
- Contactless scanning technology which can be remotely operated

Technology
- Open CV (ARUCo)
- Jupyter framework

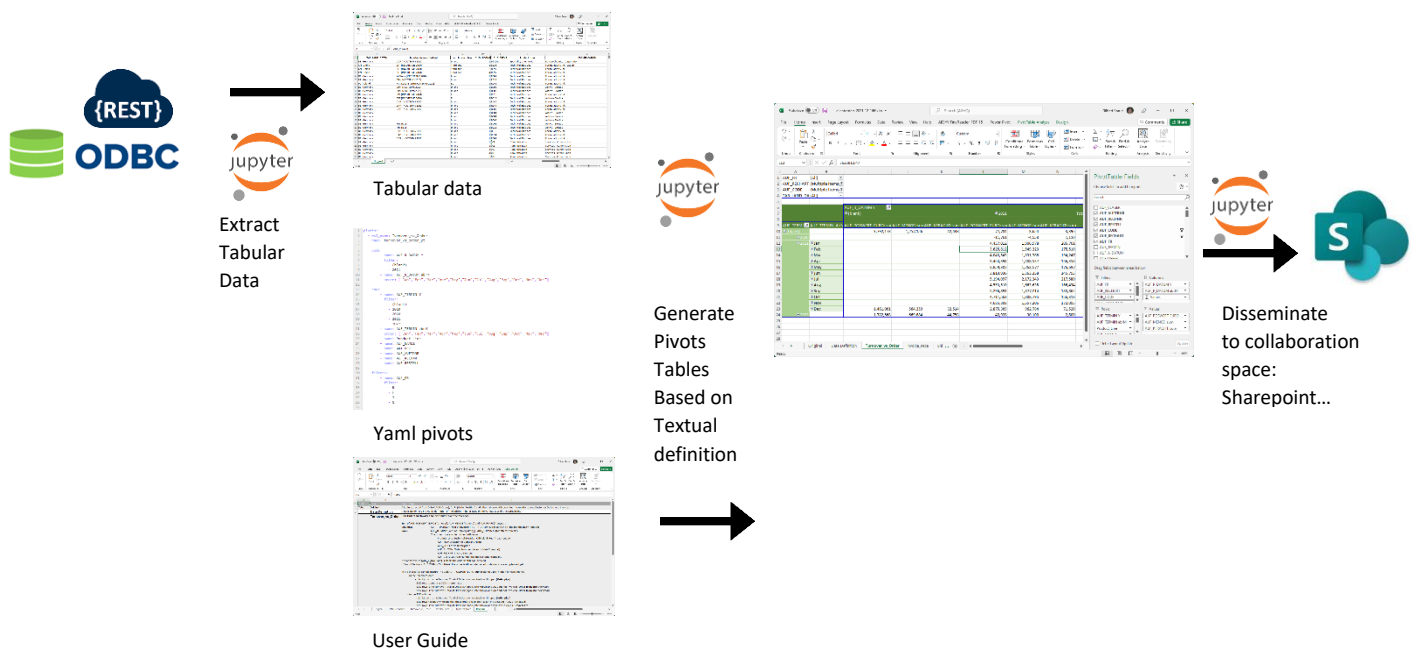## Excel Automation: database extract, pivot from yaml

Excel Table + Excel User Guide

+

Yaml pivots Definition

=

Excel File including Tabular Data, Pivots and User Guide

The problem to solve

- Company data must be analyzed, presented, and disseminated inside the enterprise.
- Nowadays, many workers are Excel savvy.
- Tabular data
  - o Is convenient to store information
  - o Is easy to retrieve from relational database systems
  - o It's not easy to read for humans
- Pivot tables are a very powerful tool to analyze and present information contained in tabular data
- But it's a tedious task to repeat this effort everyday

The solution

Using Jupyter based scripts, data is extracted from company sources into tabular format. A user-

Tabular data

Extract
Tabular
Data

Yaml pivots

Generate
Pivots
Tables
Based on
Textual
definition

User Guide

Disseminate
to collaboration
space:
Sharepoint…

guide is prepared to support user providing data definition and business process hints. A textual definition of pivots table is designed to present and drill the extracted tabular data (this textual definition enhances collaborative design with company business experts). Jupyter scripts transform all those inputs into the desired Excel file.

This information dissemination is done placing it automatically into the appropriate Collaborative spaces of the company (SharePoint etc.…)

Benefits

- Unleash the power of data disseminating it to appropriate workers, supporting them in their daily task to get more efficient and raise overall company awareness.
- Excel is available offline and doesn't require extra software licenses nor User specific training to use complex tools like ERP frontend

Technology

- Bespoke python scripts deployed in a Jupyter environment
- C# dotnet task to generate pivots based upon yaml definition file
- No Excel license required on the server side (only Office 365 or Excel license on user PC)
- Management scripts python based in the Jupyter IDE for various task
- The application is designed and deployed according to Company specific requirement using the following steps
    - Business priorities and user group identification
    - Technical definition of Company data lake access and Collaborative environment
    - Design of the tabular data, user guides and pivots specification
    - Specification of the deployment platform
    - Deployment of the solution on the platform
    - Key User training
    - On-going support for tuning the solution to business changes and solve technical issues

Business Process Management

Business Process Modelling Notation (BPMN) is a de-facto standard to capture company process nowadays.

The graphic presentation of BPMN is very compelling and quite easy to grasp with few key graphics shapes, easy and quick to learn.

It is a very good tool to share company workflow among the staff and is a real enabler for new-comers integration.
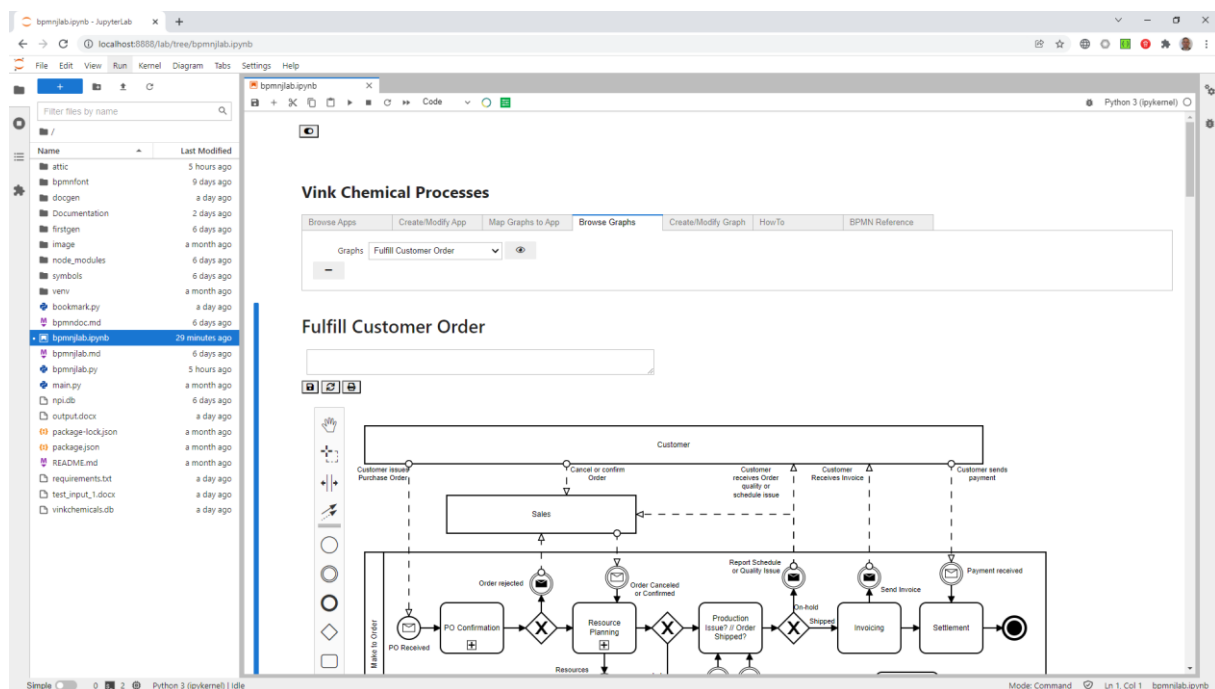
The problem to solve

While the BPMN tool is largely used, the number of graphs and their interpretation can become cumbersome as their number grow and become significant.

It is hard to maintain the correspondence between those BPMN Graphs and a textual definition of the different activities composing them.

The solution

A BPMN Graph editor is integrated to Jupyter Lab IDE allowing to organize Graphs, grouping them according to company organizational structure. For each item of the BPMN Graph, it is possible to associate a textual definition.

All this information is kept in a relational database.

Every BPMN Graph and the associated comments can be printed, like a book,

This document enables the user to see the definition of each Graph item and the little "house" next to the textual definition of the Item allow to come back to the graph. It is very easy to learn the associated Business Process. A search box enables the user to enter keywords which are then highlighted, and two arrows (up and down) enables navigating through their occurrences.

Benefits

- Single source of truth of the company Business Process
- Efficient enabler to quickly learn a company business process or for a refresh
- Time saver for the BPMN professional as it organizes efficiently collections of graphs (applications) and the textual explanation associated to them, issuing reports based on this data model capability.
- The modular underlying technical framework is highly customizable to company custom requirement

Technology

- The front-end BPMN editor and viewer comes from BPMN-IO, the Open-source flavor of Camunda BPMN German specialists.
- Jupyter Lab framework, for a very flexible integration of the editor and the Graphs and Application reporting

Annex

Jupyter

Jupyter IPython Project (2014) [Software] jupyter.org: https://jupyter.org/

Project Jupyter is a project of a community of academics and companies whose objective is to "Develop open-source software, open standards and services for interactive computing in dozens of programming languages".

Among its languages, we can mention R and Python but also Julia, C++, C#, Java...

It was derived from IPython in 2014 by Fernando Pérez (Berkeley Institute for Data Science).

One of the objectives of the community is to support scientific publication reproducibility[3] thanks to "notebooks".

The IPython notebook, now known as the Jupyter notebook is an interactive computing environment, where it is possible to combine code execution, rich text, math, plots, and rich media.

It is therefore possible to communicate in clear language, and at the same time, to propose the algorithmic resolution of the problem exposed, which can be "executed" step by step.

It is also possible to reuse many code libraries to facilitate the study of specialized scientific fields.

Thanks to web technologies, it is also possible to access shared data, which simplifies their reuse, avoids errors, and allows their distribution.

This project has received the support of many donor organizations in recent years and has benefited from the injection of tens of millions of dollars to improve its design.

Jupyter Institutional partners

- Amazon Web Services (Brian Granger, Steven Silvester)
- Bloomberg (Jason Grout, Paul Ivanov)
- Cal Poly (Ana Ruvalcaba)
- Netflix (Kyle Kelley, M Pacer)
- QuantStack (Sylvain Corlay)
- QuanSight (Matthias Bussonnier, Carol Willing)
- Two Sigma (Afshin Darian)
- UC Berkeley (Fernando Perez)
- University of Michigan (Damian Avila)
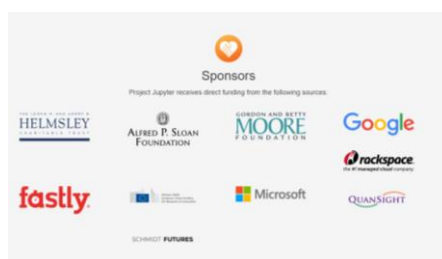- Thorn (Peter Parente)

Sponsors

Figure 6: Jupyter project donors

---

[3] Reproducible science is when anyone (including others and your future self) can understand and replicate the steps of an analysis, applied to the same or even new data.