







Data Ingestion and Inspection for Smart City Applications

P. Bellini, D. Bologna, Q. Han, P. Nesi, G. Pantaleo, M. Paolucci Department of Information Engineering, DISIT, University of Florence, Italy Department of Computer Science, Colorado School of Mines, Golden, CO 80401 USA











Context and Problems

- Smart city context includes solutions that presents
 - Data: Heterogeneous, large volume, several protocols, legacy systems, semantics, real time, multiple domains
 - Processes: several of different kinds, aperiodic, periodic, event driven,.
 - Relationships: among data and processes and mixt
 - Non Func. Req: security, GDPR, reliability, quality, scalability, etc.
 - Interoperability: legacy, protocols, modularity,

- Data Ingestion: the models and mechanism for data gathering
- Data Inspection: the model and solution to identify problems, and understand solution













Open Data:

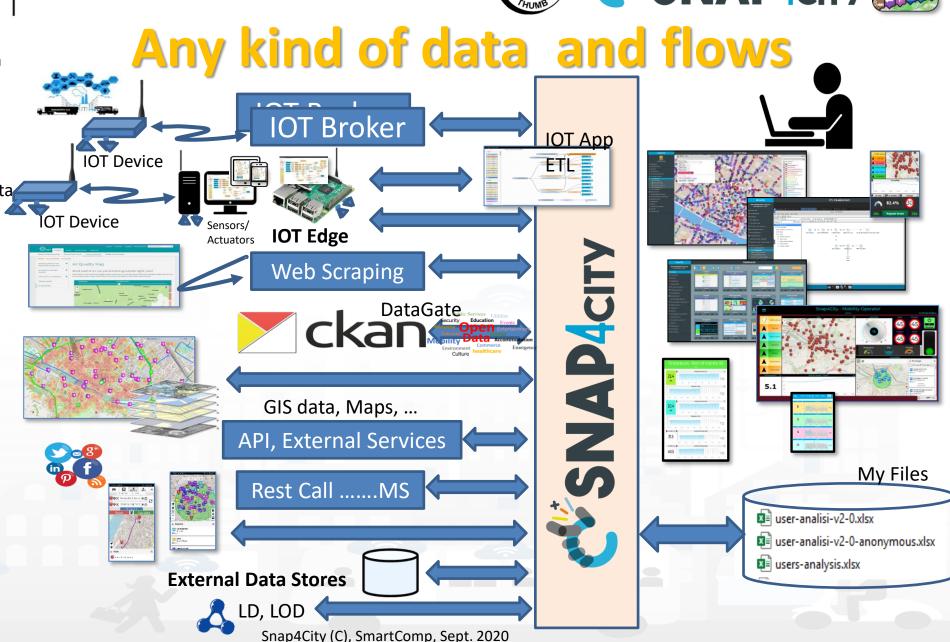
- Data gate, federation of Open **Data Portals**
- ETL processes (PULL)
- **IOT Application processes**

IOT Networks:

- IOT Application processes, data driven or PULL
- IOT Brokers (Push) → IOT Shadow

Web Pages:

- Web scraping, crawling processes
- Social media: Twitter, Facebook,...
 - Twitter Vigilance, IOT App
- **Mobile Apps**
 - Smart City API
- Files upload: CSV, Excel, etc.
 - IOT Applications, ETL
- REST API, WS, FTP, LD, LOD, etc.
 - IOT Applications, ETL
- Data base accesses
 - GIS: WFS, WMS
 - ETL, IOT Application







URBAN PLATFORM: SMART CITY IOT AS A SERVICE AND ON PREMISE



- LOCAL GOVERN
- STAKEHOLDERS
- CITY USERS
- IN-HOUSE
- ENERGY OPERATORS
- MOBILITY OPERATORS
- COMMERCIAL **OPERATORS**
- SECURITY OPERATORS
- INDUSTRIES
- RESEARCHERS
- START-UPS
- ASSOCIATIONS



- ASSESSMENT
- AUDITING

- OPEN IOT DEVICES
- IOT EDGE
- IOT GATEWAY
- PAX COUNTERS
- IOT BUTTONS
- TEST CASES, SCENARIOUS, VIDEOS, HACKATHONS
- OPEN SOURCES, COMMUNITY OF CITIES
- TRAINING TUTORIALS, COMMUNITY MANAGEMENT

IOT APPLICATIONS - INSTANT APPS



DATA DRIVEN APPLICATIONS • REAL TIME PROCESSING . BATCH PROCESSING . ANY PROTOCOL & FORMAT

DASHBOARDS & APPLICATIONS



CONTROL ROOM • SITUATION ROOM • OPERATOR DASHBORDS • BUSINESS INTELLIGENCE • WHAT-IF ANALYSIS • DECISION SUPPORT • SIMULATIONS • RISK ANALYSIS • **RESILIENCE ANALYSIS**

MOBILE & WEB APPLICATIONS



DEVELOPMENT KIT • SUGGESTIONS • MOBILE APPS MONITORING PANELS
PLATFORM UTILITIES READY TO USE SMART APPLICATIONS

MICROSERVICES & ADVANCED SMART CITY API

LIVING LAB - DEV TOOLS - COWORKING

BIG DATA - DATA ANALYTICS

DATA ANALYTICS TOOLS - MICRO-APPLICATIONS



IOT DIRECTORY • SERVICE MAP • **RESOURCE MANAGER • DATA GATE •** R STUDIO • ETL



PREDICTIONS • ANOMALY DETECTION • WHAT-IF ANALYSIS • TRAFFIC FLOW RECONSTRUCTION • ORIGIN-DESTINATION MATRICES • SOCIAL MEDIA ANALYSIS • OFFER VS DEMAND ANALYSIS • ENVIRONMENTAL DATA ANALYSIS AND PREDICTIONS • REAL TIME HEATMAPS • ROUTING • ALERTING • EARLY WARNING • PERSONAL AND VIRTUAL

KM4CITY DATA AGGREGAT KNOWLEDGE BASE - EXPERT SYSTEM OF THE CITY - BIG DATA STORE

IOT MNG - DATA MNG - DATA INSPECTOR - PROCESS MNG - USER ENGAGEMENT - GDPR MNG ...

GIS

CITY UTILITIES

OPEN DATA

LEGACY & EXTERNAL SERVICES

PERSONAL DATA

IOT / IOE

BROKERS

ASSISSTANTS • SMART SOLUTIONS • SMART SHARING • PARTECIPATORY

INDUSTRY 4.0

SOCIAL MEDIA



































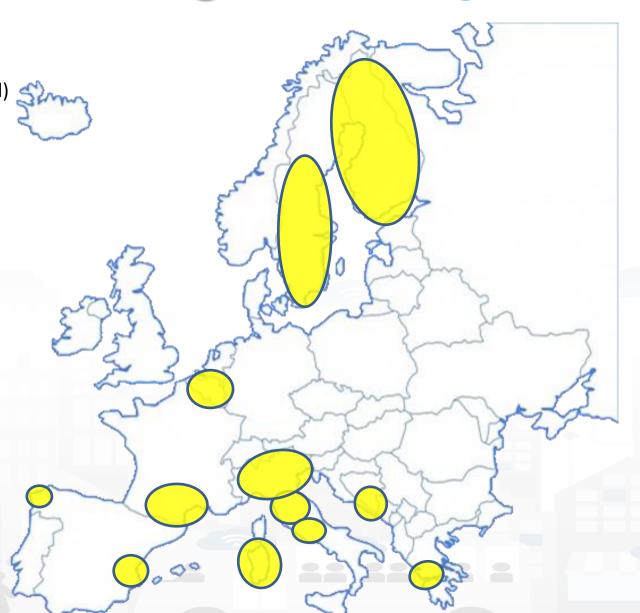


Coverage 2020



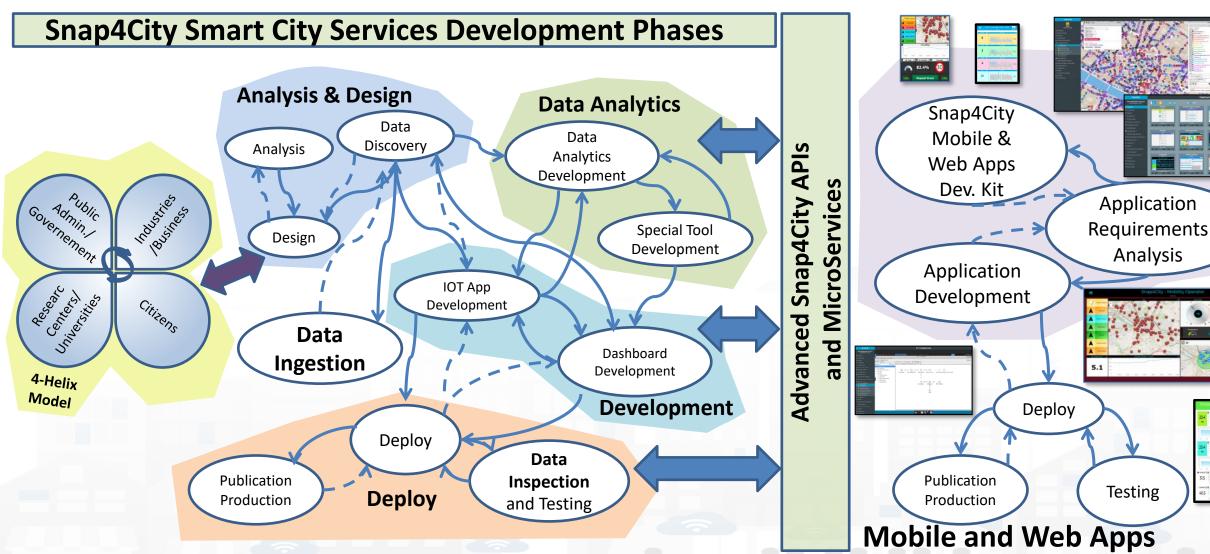
Main Organizations/areas

- Antwerp area (Be)
- Capelon (Sweden: Västerås, Eskilstuna, Karlstad)
- <u>DISIT demo (multiple)</u>
- Dubrovnik, Croatia
- Firenze area (I)
- Garda Lake area (I)
- Helsinki area (Fin)
- Livorno area (I)
- Lonato del Garda (I)
- Modena (I)
- Mostar, Bosnia-Herzegovina
- Pisa area (I)
- Pont du Gard, Occitanie (Fr)
- Roma (I)
- Santiago de Compostela (S)
- Sardegna Region (I)
- SmartBed (multiple)
- Toscana Region (I), SM
- Valencia (S)
- Venezia area (I)
- WestGreece area (Gr)













Data

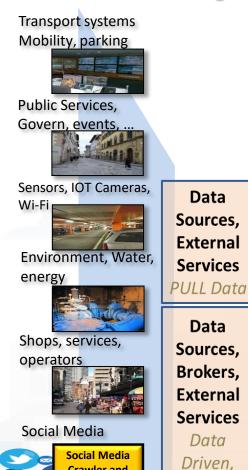
Data

Data



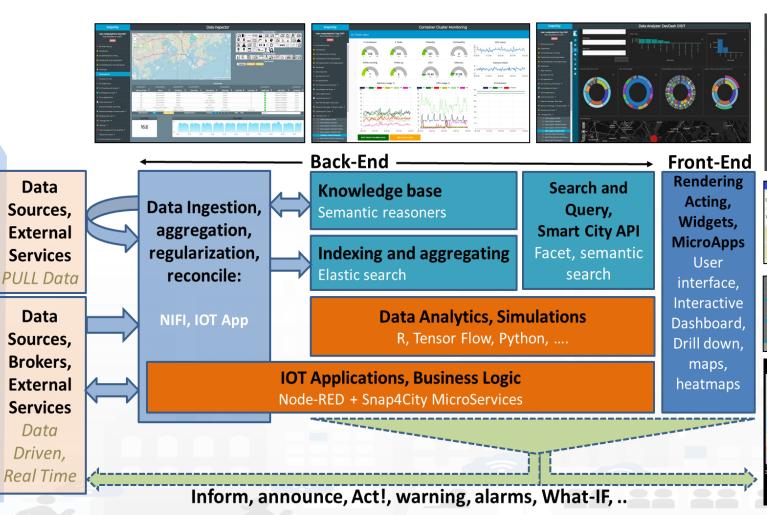


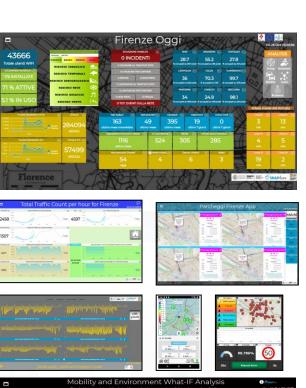
Smart City Functional Architecture



Crawler and

Manager









Data are not so Simple as one can imagine

- Data to be managed into the Smart City IOT are not so simple as one may imagine, and not limited to take into account only IOT Devices.
- THUS, a large number of data TYPES and sources have also to be addressed:
 - E.g.: external services, heatmaps, trajectories, maps, OD matrices, actuators, personal data, KPI, API descriptors, special widgets, events, predictions, Tweets, posts, GIS, mobile devices, etc.
 - With their complexity of managing data, licensing, etc...
- THEY are called High Level Types: how and which tool / process can cope with them into the Smart City Platform?... See next!
 - Thus a unified model is needed.
- In 2nd part we have shown how to show specific types of data









Unified Data and Services Model/Classification











All of them can be shown on Dashboards, what about manipulate them!!!!

HLT, High Level Types++	GPS	Static	Dynamic	Single	Time Series	Trajectory	НТТР	How to ingest/change/manage/see	
POI (Point of Interest)	Yes	Yes		Yes				DataGate, ETL, IOT App, API	
MyPOI data	Yes	Yes		Yes				Dashboard, IOT App, UserInterface, API	
KPI (metrics) data		Yes	Yes	Yes	Yes			Dashboard, IOT App, API, SQL calls x Metrics	
Dashboard-IOT App (msg)		Yes	Yes	Yes	Yes			Dashboard, IOT App, API	
Dashboard-IOT App real time			Yes	Yes	WS			Dashboard, IOT App, API WS	
My Personal Data		Yes	Yes	Yes	Yes			Dashboard, IOT App, UserInterface, API	
MyKPI data	Yes	Yes	Yes	Yes	Yes	Yes		Dashboard, IOT App, UserInterface, API	
Sensor data	Yes	Yes	Yes	Yes	Yes			Dashboard, IOT Directory, IOT App, UserInterface, API	
Sensor Actuator data	Yes	Yes	Yes	Yes	Yes			Dashboard, IOT App, UserInterface, API	
Synoptics data			Yes	Yes	Yes			Dashboard, IOT App, UserInterface	

Special Widget (complex)		Yes	Yes	Yes	Yes		Yes	Dashboard, ETL, special, IOT App, API
Complex Event (msg)	Yes	Yes	Yes	Yes	Yes			Dashboard, ETL, special, IOT App, API
WFS/WMS (GIS data)	Yes	[yes]	[yes]				Yes	Dashboard, GIS tools, or GeoServer, IOT App
GTFS	Yes	Yes	Yes	Yes	Yes			ETL, special [IOT App], MicroApplications
OD Matrices	Yes	Yes	Yes	Yes	Yes			Special tools, MicroApplications
Snap4City (C), SmartComp, Sept. 2020								







How to Ingest





All of them can be shown on Dashboards, what about manipulate them!!!!

HLT, High Level Types++	GPS	Static	Dynamic	Single	Time Series	Trajectory	НТТР	How to ingest/change/manage	
API (Ext. Srv., any prot.)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	ETL, Special, IOT App,	
External Service (web pag)		Yes					Yes	ETL, Special, IOT App, Web Scraper,	
MicroApplication (webapp)		Yes					Yes	Dashboard, IOT App, API, FTP,	
Heatmap matrix	Yes	Yes	Yes	Yes	Yes			Maps, IOT App, MicroService, UserInterface, API	
Synoptics (group)		Yes	(Yes)	(Yes)	(Yes)		Yes	Dashboard, Special Tools, IOT App, API,	
Special Tools (functional)	(Yes)		(Yes)	(Yes)		(Yes)	Yes	As MyPersonalData,	
Typical Trends (not yet)	(yes)	Yes	(Yes)		Yes		Yes	MicroApp, Special tools, (API),	

Non HLT

Traffic Flows (are coming)	(yes)	Yes	Yes		Yes	Yes	Yes	Maps, Special tools, API,
Color Maps		Yes		Yes				Maps, Tables, Special tool, User Interface, API
GTFS (see Sensors, POI)	Yes	Yes	Yes		Yes	Yes		Maps, Special tools, API,
Typical Trajectory (MyKPI)	Yes	Yes	Yes		Yes	Yes		Maps, Special tools, API,

Now, it is more clear about what we intend as:











Snap4City Portal and Integrated tools

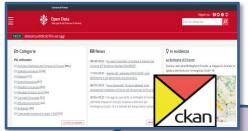


IoT App – Automatize:

- Import data from CKAN to Snap4City
- Upload Public Data from Snap4City to CKAN
- Data Harvesting
- Dashboards and Mobile/Web Apps creation



IoT App





Open or Private External CKAN Data Portals



Advanced Snap4City APIs and MicroServices

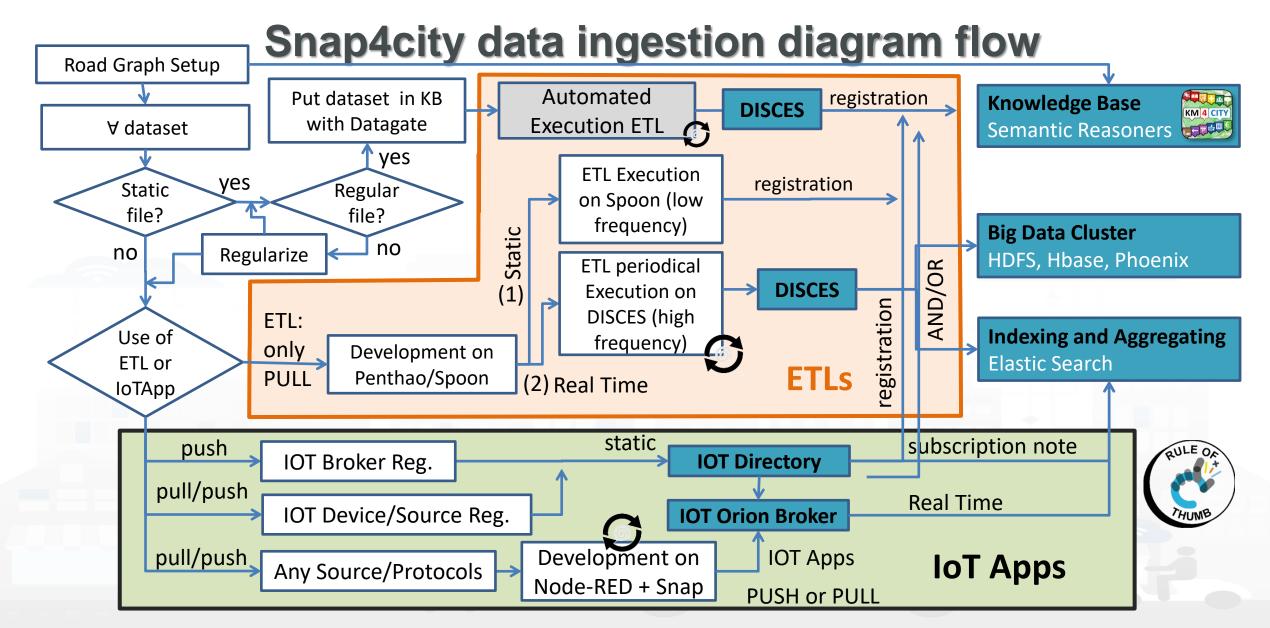


Knowledge and Storage Data from the Field and City













IOT App vs ETL



IOT Applications

- Created on browser
- A huge number of Protocols <u>https://www.snap4city.org/65</u>
- Scheduled internally and managed as Container
- Largely diffused approach as Node-RED
- Large number of Snap4City MicroServices, well documented
- PUSH and PULL models
- Simple mechanism to add new Features
- Very diffused in IOT
- Static and Dynamic data models depending on IOT Broker capabilities
- Scalable on Cloud
- Also present in IOT Edge devices

ETL processes

- Created with Spoon editor on VM (on premise or via remote access to VM)
- A Large number of protocols <u>https://www.snap4city.org/65</u>
- Scheduled by DISCES in the back office
- Well known data warehouse model
- Well documented for the process
- Only PUSH models
- Complex mechanism to add a new functionality
- Very diffused in Data transformation
- Static and Dynamic data models well linked ..









Data Inspector (Digital Twin info) Major Submodels

Digital Twin

- Device and sensors data
- Values
- Healthiness criteria and values
 - Machine learning tools
- Images and phyiscal world
- Licensing
- -Users

Process Views

- Device Management tool
- Data ingestion processes
 - ETL, IOT Apps
- Data storage access views
 - Index views
 - Relatioships view
- Data Analytics and Trasformation
 - IOT App, R Studio, Python
- Data Rendering Dashboards
- Processes' Developers



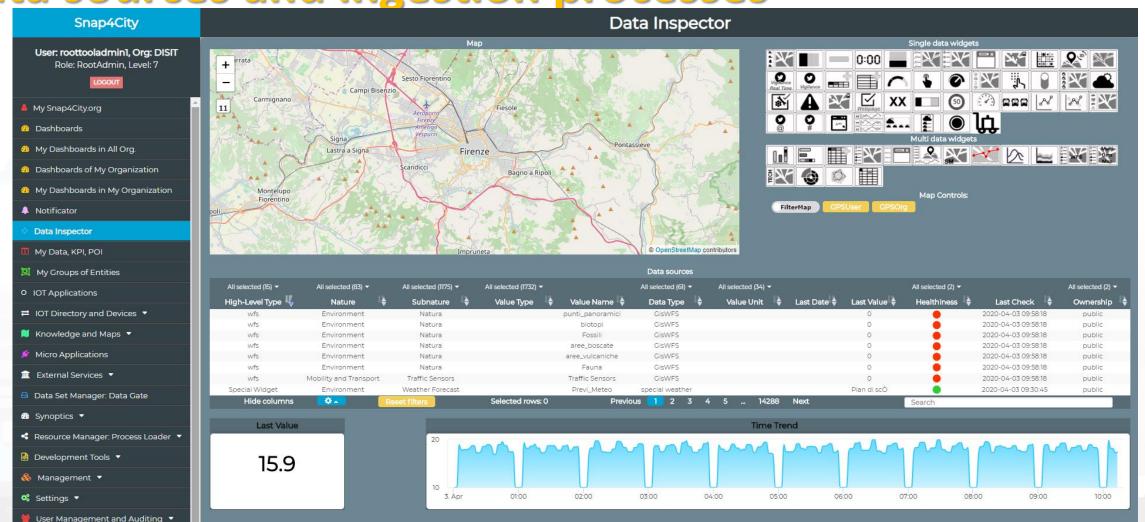






Data Inspector: all you need to know about data,

data sources and ingestion processes

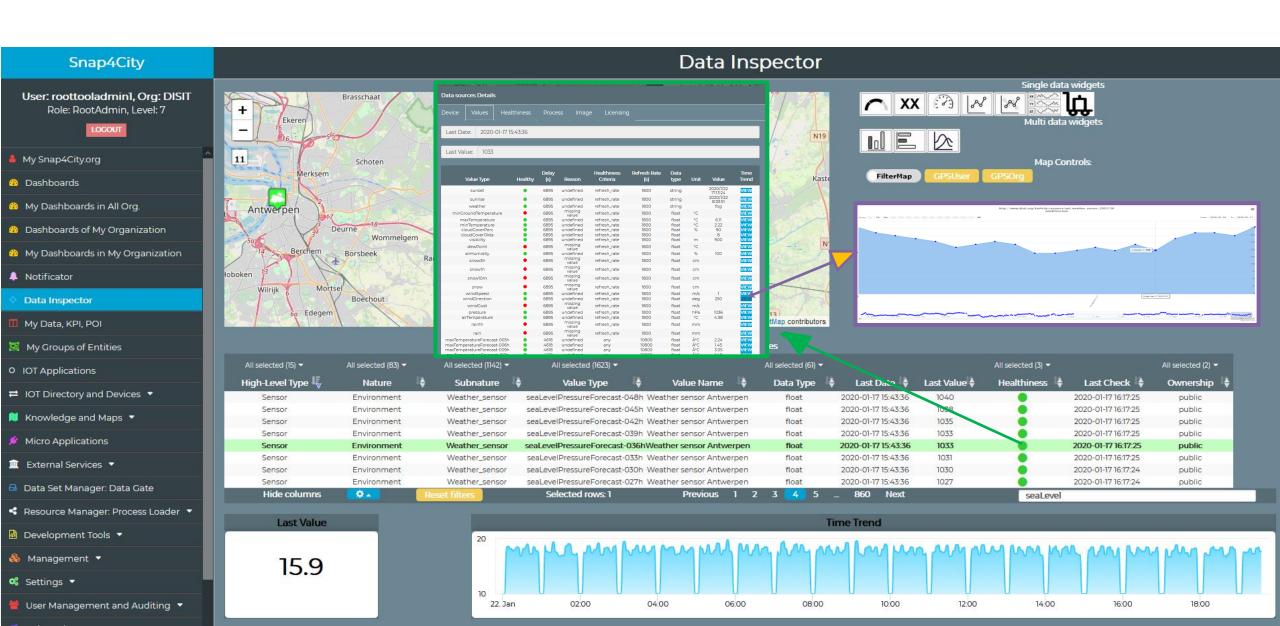


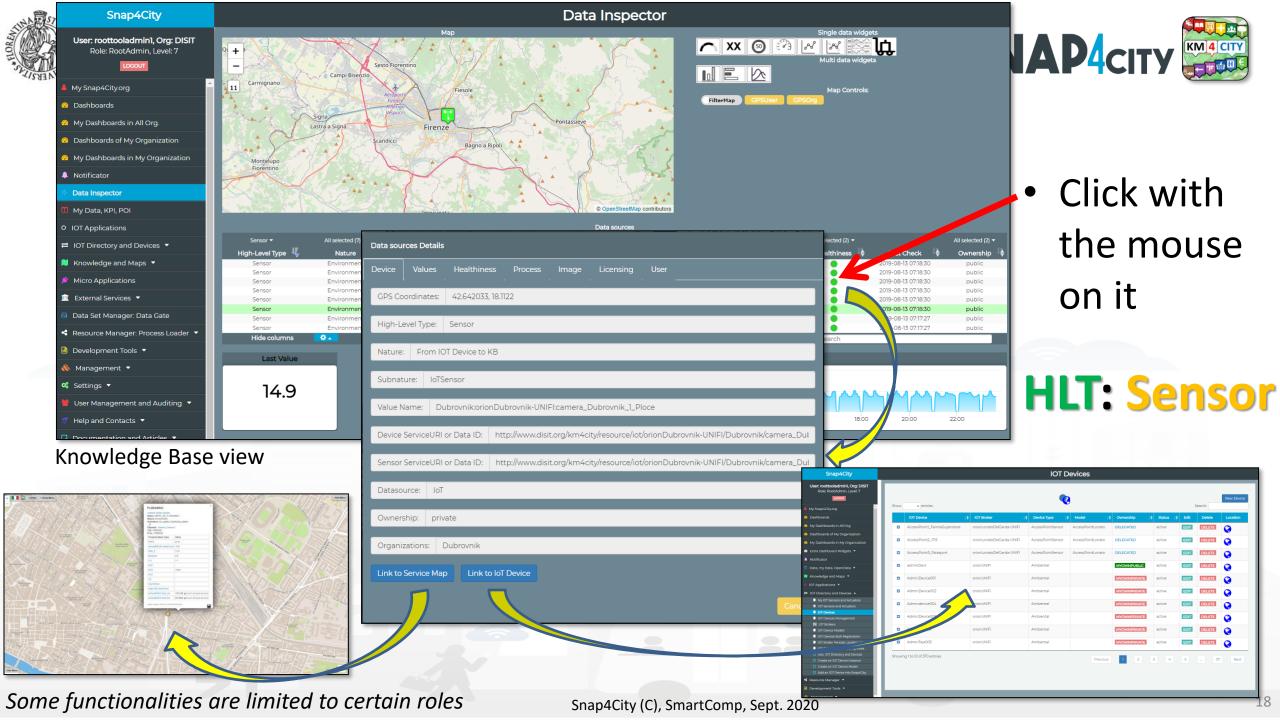












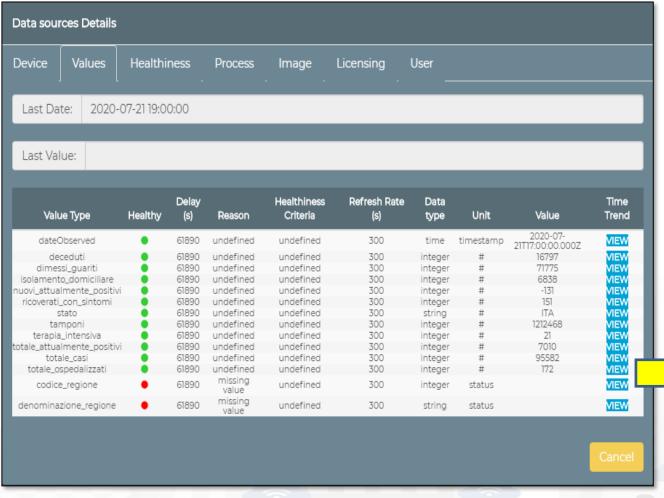




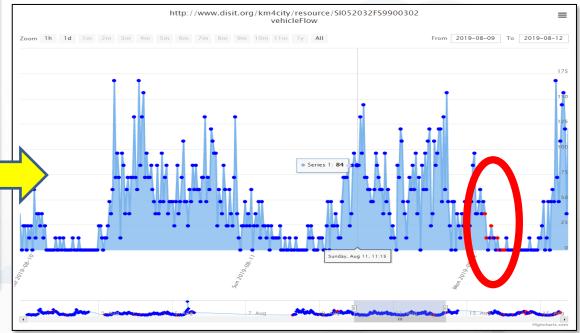




HLT: Sensor



- Specific values of selected
- Information of the values of the other sensors on the same device
- View Trends, marking problems, healthiness by point according to a Fuzzy model
- Marking problems for future machine learning processes (separate tool)



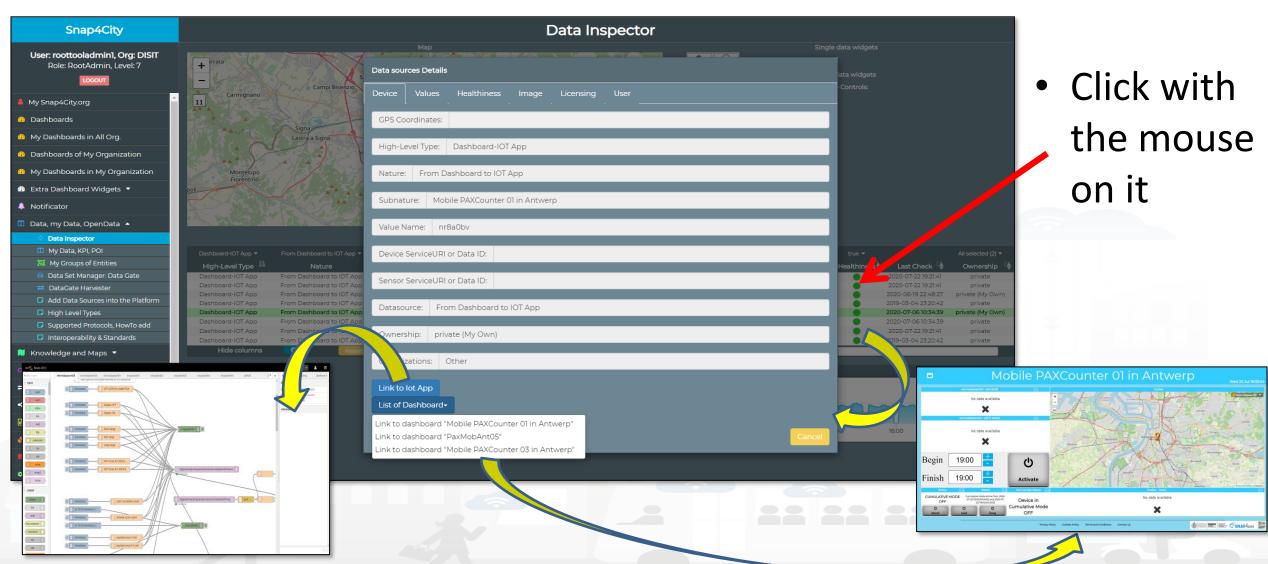








HLT: From Dashboard to IOT APP



Snap4City (C), SmartComp, Sept. 2020

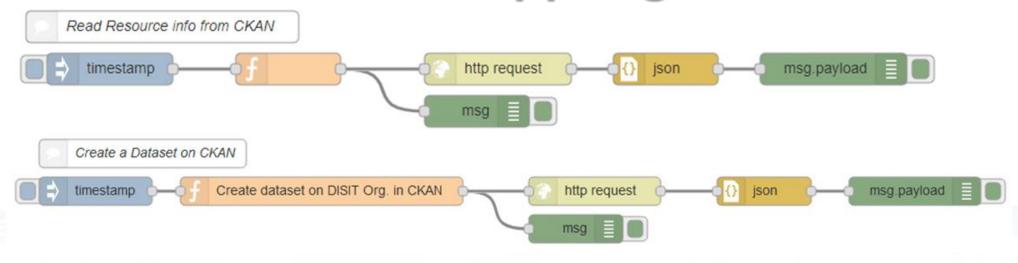




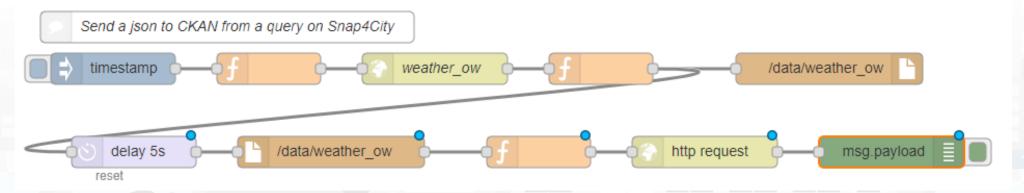




Some IOT App segements



Almost all the calls to CKAN are quite similar

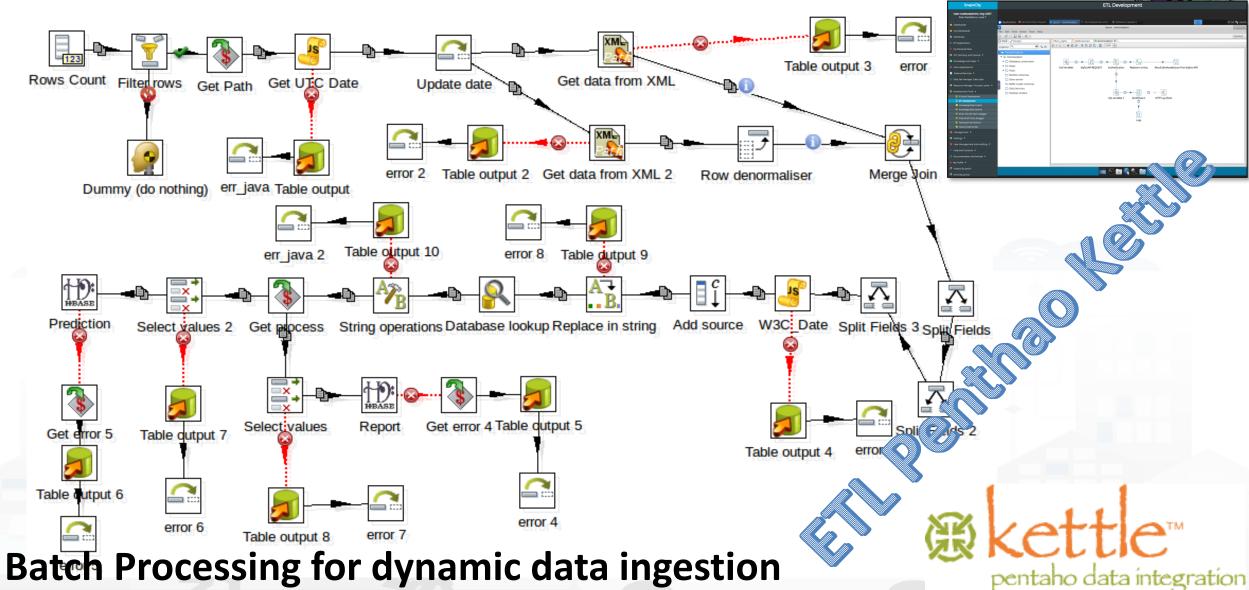








Integrated The Avelopm













	Datagate	ETL	IoTApp		
types of data	S	S, P	S, P, RT		
managed					
Data protocol	PULL	PULL	PULL and		
types managed			PUSH		
Scheduling	external	external	internal		
Flows to manage	N	N	1		
N instances of					
the same dataset					
Users' technical	without	medium/high	low		
level					
Development	1,2 hours	1, 2 weeks	3, 4 days		
time					
Semantic	standard	ad hoc (man-	ad hoc (semi-		
(KM4City)	template	ual)	automatic)		
Developed num-	1334	162	76		
ber	datasets				
Mean number of	0	120.333	27,67		
blocks					
Mean number of	0	275	229		
lines of code					





Conclusions



Problems

 Complex data models, multiple processes, multiple tenancy/organizations, etc.

Integrated approach from ingestion and inspection

- Formal methodology support:
 - data discovery, development, living lab
- Unified data model, avoiding pillars since the data model
- Formal Model support:
 - data vs processes vs developers/owners
- Powerful Tool as Data Inspector exploiting knowledge base as expert systems: data, processes, relationships, events, etc.

Acknowledgements

- Thanks to the European Commission for founding. All slides reporting logo of Snap4City https://www.snap4city.org of Select4Cities H2020 are representing tools and research founded by European Commission for the Select4Cities project. Select4Cities has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation Programme (grant agreement n° 688196)
- TRAFAIR is a CEF project. All slides reporting logo of TRAFAIR project are representing tools and research founded by the EC on CEF programme http://trafair.eu/
- Thanks to the European Commission for founding. All slides reporting logo of REPLICATE H2020 are representing tools and research founded by European Commission for the REPLICATE project. **REPLICATE** has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation Programme (grant agreement n° 691735).
- Thanks to the European Commission for founding. All slides reporting logo of **RESOLUTE H2020** are representing tools and research founded by European Commission for the RESOLUTE project. **RESOLUTE** has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation Programme (grant agreement n° 653460).
- Thanks to the MIUR for co-founding and to the University of Florence and companies involved. All slides reporting logo of Sii-**Mobility** are representing tools and research founded by MIUR for the Sii-Mobility SCN MIUR project.
- **Km4City** is an open technology and research line of DISIT Lab exploited by a number of projects. Some of the innovative solutions and research issues developed into projects are also compliant and contributing to the Km4City approach and thus are released as open sources and are interoperable, scalable, modular, standard compliant, etc.











INEA CEF-TELECOM Project funded by European Union





European Union Funding for Research & Innovation























GREEN FIELD PEAS





















CONTACT

DISIT Lab, DINFO: Department of Information Engineering Università degli Studi di Firenze - School of Engineering

Via S. Marta, 3 - 50139 Firenze, ITALY https://www.disit.org









Email: snap4city@disit.org

Office: +39-055-2758-515 / 517

Cell: +39-335-566-86-74 Fax.: +39-055-2758570

