FP7 Project ASAP
Adaptable Scalable Analytics Platform

D10.7 Showcase

WP 10 - Exploitation and Dissemination
ASAP Consortium

Nature: Report
Dissemination: Public

Version History

Version	Date	Author	Comments
0.1	06 Feb 2017	A. Scharl	Initial Version
0.2x	Feb 2017	WP Leaders	Partner Contributions
0.3	24 Feb 2017	P. Pratikakis	Workflow Management Section
0.4	06 Mar 2017	W. Rafelsberger	Minor Revisions
1.0	08 Mar 2017	A. Scharl	Revision and Final Edits

Acknowledgement

This project has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under grant agreement number 619706.

Executive Summary

This showcase deliverable D10.7 presents a summary of the goals and results of the ASAP project (“Adaptable Scalable Analytics Platform”). It represents a compact overview document that (i) outlines the workflow management tools within the ASAP platform, (ii) presents the application of the platform in conjunction with two use cases on Web content and telecommunications data analytics, including screenshots of the ASAP dashboard to render the resulting datasets, (iii) provides a list of links to the relevant software and data resources, including the project’s GitHub repository with detailed licensing information, and (iv) concludes with the profiles of the partner organizations and the contact information of participating researchers.

Table of Contents

1. Introduction and Project Goals

The ASAP FP7 research project[1] developed a dynamic open-source execution framework for scalable data analytics. The underlying idea is that no single execution model is suitable for all types of tasks, and no single data model (or store) is suitable for all types of data. Complex analytical tasks in multi-engine environments therefore require integrated profiling, modeling, planning and scheduling functions.

The ASAP platform facilitates managing such complex analytics applications by providing a common way to construct, manage, maintain, and execute workflows. To support the building and execution of complex analytics workflows, the ASAP project had four main goals:

- A generic *task-parallel programming model* in conjunction with two runtime systems for distributed or parallel execution in the cloud. The runtimes include state-of-the-art features such as
  - irregular general-purpose computations,
  - resource elasticity and synchronization,
  - data-transfer, locality and scheduling abstraction,
  - the ability to handle large sets of irregularly distributed data, and
  - fault-tolerance.
- A *modeling framework* that constantly evaluates the cost, quality and performance of available computational resources in order to decide on the most advantageous store, indexing and execution pattern.
- An *adaptation methodology* to enable analytics experts to amend submitted workflows by changing or modifying a workflow while it is being processed. Users can change the parameters of operators already comprised in the workflow, or the structure of the workflow by removing or adding operators.
- A *visual analytics dashboard*[2] to show query results and metadata in an intuitive manner, with special focus on the interactive exploration of datasets, dynamic temporal controls, on-the-fly query refinement mechanisms, and the geospatial projection of structured and unstructured data.

2. Managing Analytics Workflows with the ASAP Platform

Complex analytics applications combine multiple operators that process data in various formats, possibly from multiple data stores and data sources. The ASAP platform can be used to support the collaboration between:

1 www.asap-fp7.eu
2 asap.weblyzard.com

- *Operator Developers:* Expert programmers that design and implement analytics operators.
- *Workflow Designers:* Domain experts that combine operators and data sources to construct complex workflows.
- *Users:* Non-expert workflow users, such as marketing experts trying to discover trends media coverage or customer data.

ASAP connects these roles by providing a common platform for the management of analytics workflows at all these levels of abstraction. Deliverable D1.3 describes the final ASAP platform architecture design, defining an extensive set of user requirements and use cases. Deliverables D1.1 and D1.2 present earlier versions. In terms of platform development, ASAP was implemented by the project partners by means of a portfolio of interconnected components:

1. A *Workflow Management Tool* (WMT) that allows the analytics user to see high-level descriptions of operators and data stores, design abstract workflows, and optimize them. Deliverable D5.2 describes the WMT, built following the design presented in Deliverable D5.1.
2. A new *Operator Definition Language* that help developers express irregular operators, not supported by previously existing analytics engines. Deliverables D2.1 and D2.2 present the design of an operator language for irregular applications.
3. An *Intelligent Resource Scheduling* (IReS) platform that profiles operators, learns their cost, selects optimal plans for workflow materialization, and executes them. While Deliverable D3.1 outlines the design of the IReS platform, the subsequent Deliverables D3.2 and D3.3 describe the first and second version of the IReS platform, respectively.
4. An *Analytics Engine* that is able to execute recursively parallel computations. Deliverable D4.1 presents the design and Deliverables D4.2 and D4.3 present the first and section version, of the Spark-Nesting execution engine.
5. An *Online Monitoring and Recalibration Platform* that enables workflow designers to adapt executing workflows. Deliverables D5.3 and D5.4 describe the support for online monitoring and adaptation of executing workflows.
6. Information Visualization (InfoViz) services presenting analytic results to the users. Deliverable D6.1 summarizes the design, Deliverables D6.2 and D6.3 describes the 1st and 2nd version of the InfoViz Services. Deliverable D6.4 presents the InfoViz Services deployed into an integrated visual analytics dashboard. Deliverable D6.5 presents the results of the usability evaluation.

Virtual Machine (VM) Deployment

We have integrated all the ASAP modules, together with several state-of-the-art analytics frameworks into a set of virtual machine images that can be used to rapidly deploy ASAP on a datacenter or on cloud nodes. The tools integrated include the IReS workflow scheduler, the web-based UI workflow management tool, Spark-Nesting (with support for recursive UDFs), the Swan single-node scheduler, as well as a battery of languages and tools, such as full python, open-source scientific and machine-learning libraries, standard Spark, Hadoop and HDFS, and PostgreSQL.

The VMs also include an array of analytics operators (often implemented by the integrated runtimes and libraries) integrated into all relevant ASAP modules, so that a prototype workflow can easily be constructed and scheduled using the web-based UI of the ASAP workflow management tool. Deliverables D7.1, D7.2, and D7.3 describe the integration process, methodology, and results.

To quickly deploy ASAP modules in the provided VMs in a new cluster, the user should boot a VM with maximum memory and cores per available node, selecting one node for the master VM. As all the ASAP software is already installed, the only configuration necessary is networking and setting the IPs correctly. The ASAP documentation[3] describes the steps required and the related configuration files and scripts. One could certainly over-provision VMs, but we have found that performance is slightly better with one VM per node.

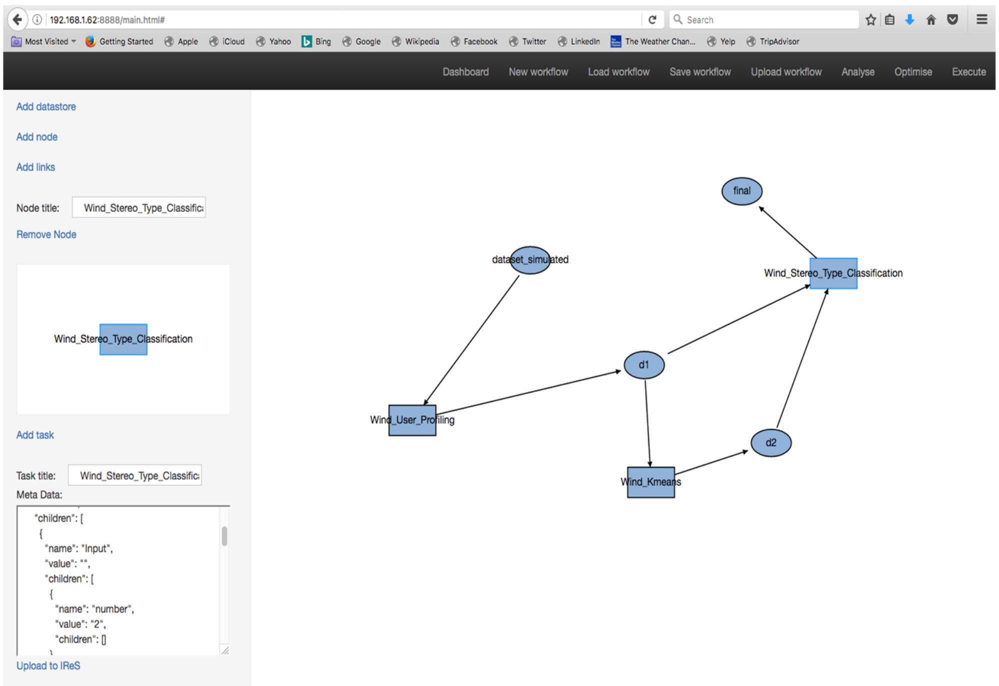

Figure 1. ASAP Workflow Management Tool

[3] project-asap.github.io/ASAP-documentation

Workflow Management

After the VMs are deployed and running, the user can connect to the Workflow Management Tool (Figure 1), by pointing a browser to the master VM. There, using drag-and-drop, the Workflow Designer can combine existing operators from the library to construct a workflow, materialize it using IReS, and execute it. The integrated ASAP documentation presents example workflows in a step-by-step guide.

Workflow Execution

By following links from the WMT UI, Users can connect to the IReS Workflow Scheduler (Figure 2), and to the data store (HDFS) and runtime engines (Spark) that execute each operator, and monitor the execution.

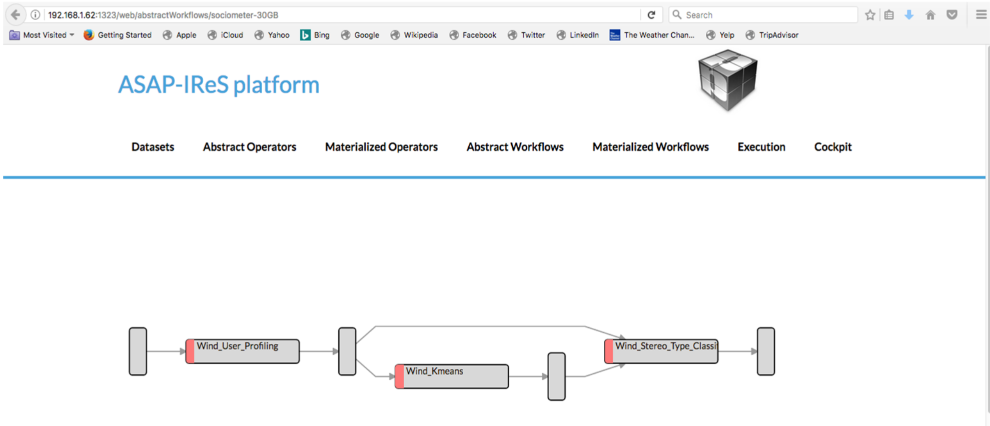

Figure 2. IReS Workflow Scheduler

Workflow Modeling

ASAP aims to help users to better manage existing workflows. To insert an existing workflow into ASAP, an Operator Developer needs to insert each operator in the IReS library (the ASAP documentation includes a step-by-step example), so that all operators appear in the WMT UI. It is then straightforward to reconstruct the new workflow and use IReS to materialize it. IReS uses profile data from the executions to learn cost models for operators with more than one alternative implementations and eventually create optimal schedules for subsequent executions.

Information Visualization

The main goal of WP6 is the development of high-performance visual interfaces as part of the ASAP system architecture to help analysts navigate big data repositories, in real time and across multiple dimensions (temporal, spatial, etc.). Special emphasis is placed on sub-second response times of the user interface, and the ability to incorporate metadata as additional context information when analyzing complex relations in Web content (WP8) and telecommunications data (WP9), enriched by geotagged news and social media content.

For effective data interchange with the other technical work packages and the flexible integration of the developed D3.js-based[4] visualizations (D6.1-D6.3) into various applications, an open Application Programming Interface (API) for structured and unstructured data was developed (D6.3). This API not only allows rendering multi-source data for the use case applications of WP8 and WP9, but also plays an important role in the ASAP exploitation strategy (D10.6). It supports a flexible *Visualization-as-a-Service* (VaaS) or, when bundled together with data acquisition and transformation services, even a *Container-as-a-Service* (CaaS) approach.

The ASAP Dashboard (D6.4)[5] integrates the developed visualizations and enables the interactive access of content feeds from different sources. The full documentation of components and available options is available online,[6] and will be actively maintained and extended beyond the lifetime of the ASAP project. The dashboard supports a rapid synchronization of multiple coordinated views, with each view representing a visual analytics component; for example:

- *Geographic Map* with (i) *custom base layers* and *data manipulation options*; (ii) incremental versions of statistical data visualizations – circular markers, choropleth; (iii) dynamic clustering to allow adaptive data exploration from high-level views to regional/local data; selected features are available under an Apache open source license.[7]

- *Charting library* with interactive features and seamless transitions, using color coding to show metadata attributes. The portfolio includes a line chart, donut chart, radar chart and scatterplot module (providing flexible rendering options for both Web content metrics and ingested statistical data).

- *Interactive temporal controls* that allow a quick selection of the analysis interval via an area chart; together with *adaptive tooltips* and *context menus*, they are an effective and intuitive way to trigger on-the-fly query refinements.

The components listed above are synchronized by means of an event and notification system for intra-module communication that supports the manipulation of intermediary and final data sets. The significant performance gains achieved are a result of improved *data structures* for document mapping, *optimized queries*, and a revised *indexing strategy*.

[4] www.d3js.org
[5] asap.weblyzard.com
[6] www.weblyzard.com/interface
[7] www.github.com/weblyzard/infovyz

3. Use Cases

The generic nature of the ASAP architecture supports a wide range of different tasks. Within the project, the consortium has focused on the real-time analysis of Web content and telecommunications data.

Use Case 1 - Web Content Analytics

The use case is centered on the information services of *Internet Memory Research* (IMR) as part of the *Mignify* platform.[8] IMR collects, cleans and classifies data from Web, and uses the results to support its online services. Deliverable D8.1 describes the initial dataset provided by IMR, Deliverable D8.2 describes the application design, Deliverable D8.3 the implementation of the workflow, and Deliverable D8.4 the results of the workflow evaluation within ASAP.

Analytic Goals

The general goal of the IMR data collection, extraction and classification processes is to build and maintain a *catalog* of product references, and to discover product offers related to this catalog on public marketplaces.

- A *Catalog* is a tree of categories and subcategories. An example of category is *Coffee machines*, and a sub-category is *Espresso machine.*
- A *Product Offer* is an online proposal to sell one or several items of a product, with specific conditions such as price, delivery, etc. If, for instance, an electronic marketplace proposes 100 items of the coffee machine xxP34, at a given price YY, this constitutes a product offer for product xxP34.

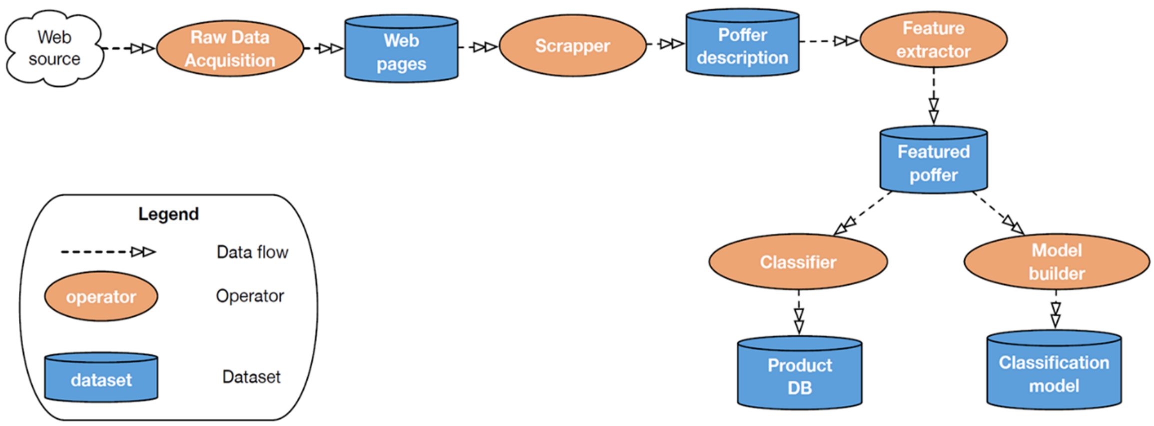

Figure 3. Workflow of WP8 – Web Content Analytics

Figure 3 shows the overall workflow of WP8. IMR maintains a *Web map* of classified sites that references hundreds of thousands of marketplaces. Our crawler scans the pages and identifies those that contain lists of products. Thanks to a semi-supervised

8 www.mignify.com

approach, we then analyze the page structure and produce a wrapper to extract a product offer record. This integrates a rich set of product-related information including brand, type, price, textual description, and user comments. Once product information has been extracted from the page, we run a classification process to predict the category of the product. The *product matching* operation associates product offers with product categories, given the description of offers extracted from e-marketplaces.

Applications

IMR maintains and expands a large database of classified product offers. This ProductDB database supports several services which can be split in two main categories depending on their target users:

- *Presence*[9] is a B2B service that takes advantage of the ProductDB database to provide competitive intelligence. The service offers a public interface that lists of eCommerce sites where a specific brand can be found. Sellers can get specific data on their brand and analyse their main competitors.
- *Bomerce*[10] is a price comparison application for Web and mobile devices. When exploring online offers for products or services, users are confronted with heterogeneous offers from proprietary eCommerce sites. *Bomerce* helps to compare such offers and seek third-party advice, sends out of notifications in the case of promotions, and checks the reputation of an eCommerce site.

Both *Presence* and *Bomerce* depend on the quality of the ProductDB database, and therefore on the data acquisition, extraction and classification workflow. We modeled and implemented this workflow with ASAP. First, an *abstract workflow* has been defined as a high-level view of the various steps involved in the transformation of raw Web pages into structured and classified product descriptions. This abstract workflow is then implemented through concrete operators taken from the ASAP library.

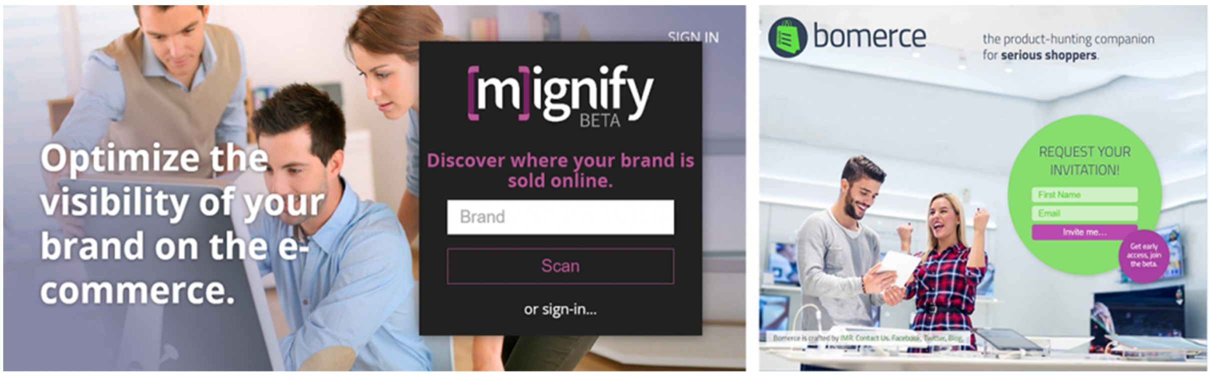

Figure 4. WP8 Applications – *Presence* and *Bomerce*

9 presence.mignify.com
10 bomerce.io

Interactive Exploration

The ASAP Dashboard shown in Figure 5 demonstrates the potential of big data technologies in conjunction with advanced visual analytics to automatically transform noisy and unstructured Web content into valuable repositories of actionable knowledge. Processing dynamic content streams from multiple sources and extracting metadata attributes from the product offers, it extends IMR's price comparisons by:

- Visualizing *aggregated keywords* computed from noisy textual descriptions contained in the product offers collected by IMR; identifying the *leading sources* of these offers, including an analysis of keywords that e-commerce sites associate with specific products or an entire product category.
- Exploring *product features* that impact the perception of a product in online media coverage (news channels vs. social media vs. product offers), creating additional value for sales and marketing decision makers.
- Providing *metadata dimensions* such as sentiment, which indicates whether a feature is mainly perceived as a *unique selling proposition* that causes satisfaction, or a *hygiene factor* that causes dissatisfaction. This distinction is an important source of feedback to guide strategic marketing decisions.

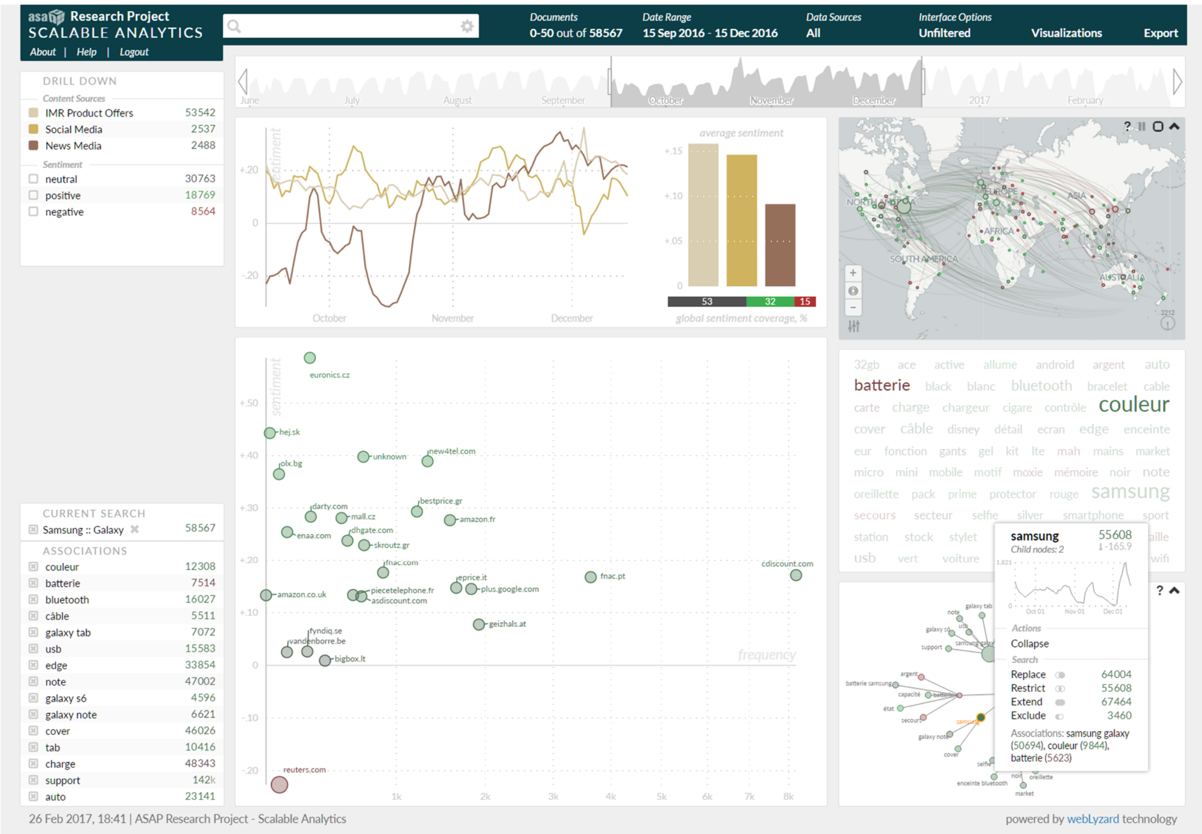

Figure 5. Screenshot of the ASAP dashboard - drill down sidebar to compare product sentiment for the Samsung Galaxy series by source; scatterplot for cross-media analysis; geospatial projection of referenced locations; tooltip for on-the-fly query refinement

For a business intelligence tool to complement price comparisons, such metadata dimensions are particularly important. The drilldown sidebar of the ASAP dashboard shown in Figure 5 helps to better understand the temporal distribution of metadata attributes. The line chart in the shown example compares the average sentiment for the *Samsung Galaxy* series by source (product offers, social media, news media). The bar chart presents the same data in aggregated form, and the scatter plot maps the frequency vs. sentiment matrix of the major content sources. The geographic map projects the entire set of search results. The adaptive tooltip in the lower right corner enables on-the-fly query refinements - either to *Replace* the search query with a new term, or to apply the Boolean operators AND (*Restrict*), OR (*Extend*) and NOT (*Exclude*). Using the dashboard's view synchronization mechanism, the tooltip is tightly coupled with the tag cloud – which highlights the product features "color*"* and *"battery"* as the strongest associations with the hovered keyword "*Samsung"*.

Use Case 2 - Telecommunications

The usage of mobile phones and the resulting datasets stored in *Call Data Records* (CDR) represent a high-quality proxy to better understand human mobility behavior in different application scenarios such as smart cities, transportation planning and environmental monitoring.

Deliverable D9.1 describes the initial WIND dataset, Deliverable D9.2 the application design, Deliverable D9.3 the first implementation of the workflow, and Deliverable D9.4 the final application and results of the workflow as evaluated within ASAP.

Analytic Goals

Mobile phones communicate to antennas covering specific local areas. The active connection between a phone and an antenna, e.g. a phone call or a text message, represents spatio-temporal information that, once collected and aggregated, reflects the distribution of users in an area covered by mobile services. The analytic goals of ASAP focused on better understanding such aggregated data - applying strict data anonymization procedures and addressing the involved computational challenges.

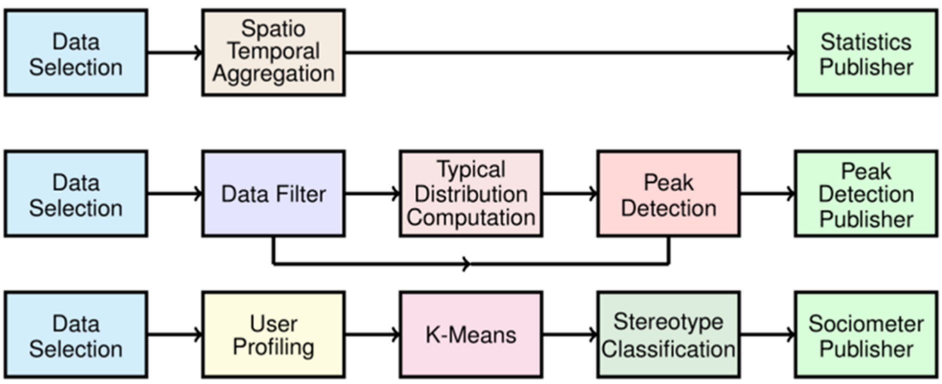

Figure 6. Workflow of WP9 – Telecommunication Data Analytics

Within ASAP, a telecommunications application was designed to demonstrate how analytical services based on human mobility data can be provided during routine mobile network operation. Figure 6 shows the workflow of selected modules of this application, including user profiling, sociometer and peak detection.

Applications

The sheer volume of CDR data poses computational challenges when collecting, storing, mining, and visualizing specific indicators. In this context, ASAP investigated the following applications (see Figure 7 for two conceptual diagrams):

- **Event Detection** analyses the different features of an event, including its spatio-temporal characteristics, social aspects, and statistical properties. By controlling input parameters such as time interval, spatial area and additional CRM attributes, analysts gain a detailed understanding of evolving events.
- **Ridesharing** provides functions for mobility managers and individual drivers alike, e.g. the visualization of routine trips in a specific area, together with an optimized car sharing solution for managing such trips. A driver can use this application as a recommender system to identify ridesharing opportunities.
- **Tourism Observation.** The analysis of dynamic tourist flows allows mobility managers to identify common movement patterns of visitors, using a map-based dashboard to provide spatio-temporal constraints as input. Along this line, mobile phone data, despite their limited spatial precision compared to other location data such as GPS, are of interest due to their global availability across countries, and their independence from specific means of transportation. We can use this kind of data to reconstruct mobility and traffic flows by using origin / destination matrices, complemented by other sources such as CDR-based traffic intensity, traffic migration from handovers, and individual movements of volunteer participants.

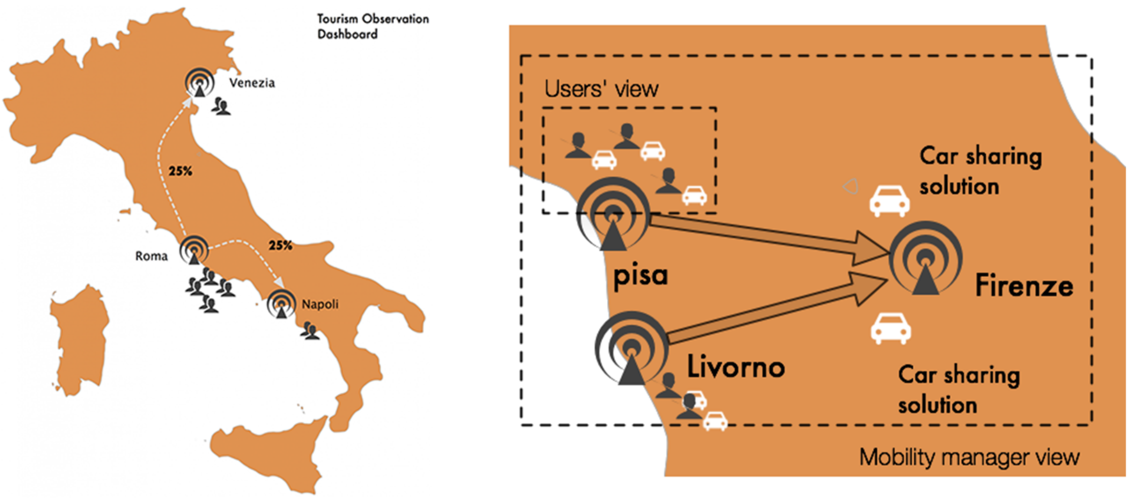

Figure 7. WP9 Applications – Tourism Observation Board and Mobility Manager

Interactive Exploration

The ASAP dashboard illustrates how actionable knowledge can be visualized from the ingested statistical indicators - to understand weekly patterns of user movement, for example, or to visualize the interplay between multiple indicators by generating an overlay of social media content on top of aggregated CDR data.

To index and visualize the statistical data produced by the ASAP workflow - e.g. area presence by user type (*resident*, *dynamic resident*, *commuter*, *visitor* or *passerby*), O/D matrices, etc. - a *Statistical Data API*[11] was developed following the *RDF Data Cube Vocabulary* approach.[12] It supports JSON to enable rapid visualization of large datasets. The desired slices of a dataset, e.g. the residents who visit a certain area, do not need to be defined within the dataset, but can be specified at runtime.

Tested with queries that delivered up to 100 million documents, the geographic map of the ASAP dashboard has been optimized for large datasets. Figure 8 shows the maximized version of the map including a tooltip for on-the-fly query refinements.

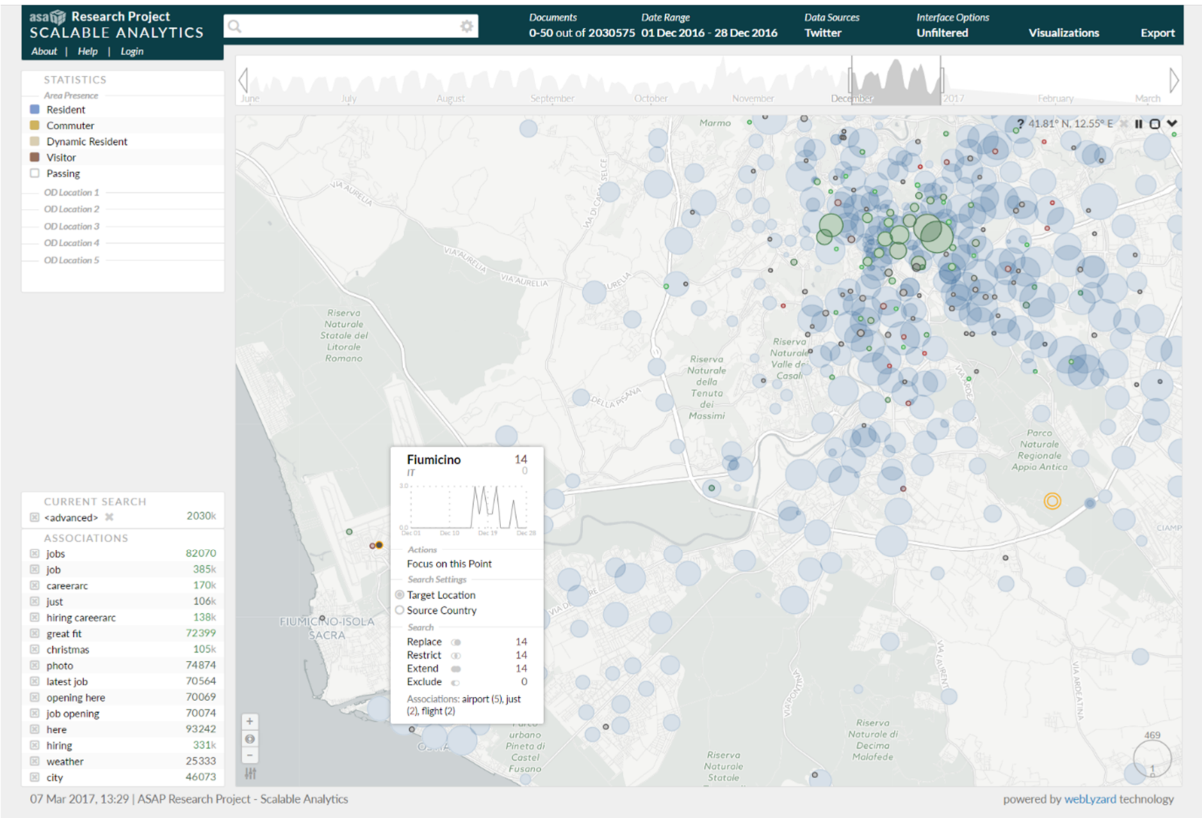

Figure 8. Geographic map of the ASAP dashboard, showing anonymized *Call Data Records* (blue markers) for Rome, and an overlay of sentiment-annotated Twitter content (green = positive; red = negative coverage) as of December 2016

[11] api.weblyzard.com/doc/ui/#/Statistical_Data_API
[12] www.w3.org/TR/vocab-data-cube

Users can zoom to street level and visualize anonymized cell tower activity data from the City of Rome (blue markers), and overlay this information with geotagged Twitter postings (green = positive sentiment; red = negative sentiment) – e.g., to identify communication hotspots during an event. The examples are intended as a proof of concept how semantic technologies in conjunction with advanced visual tools can transform statistical data into valuable repositories of actionable knowledge.[13]

4. Software and Data Resources

Prototypes have been made publicly available for download; the project Website provides links to the **GitHub.com** public repository at www.github.com/project-asap, and lists the currently available components:

- **ASAP Source Code Documentation**
  project-asap.github.io/ASAP-documentation

- **Intelligent Resource Scheduler (IReS)**
  The open source platform for optimizing, planning and executing complex analytics workflows in multi-engine environments.
  www.github.com/project-asap/IReS-Platform
- **Telecom Analytics**
  Spark Application for peak detection in call detail records
  www.github.com/project-asap/telecom-analytics
- **Platform for Analytics Workflows**
  Unified execution framework for scalable data analytics to facilitate the execution of general-purpose analytics queries over irregular data
  www.github.com/project-asap/workflow
- **ASAP Operators**
  Operator definitions for ASAP
  www.github.com/project-asap/asap_operators
- **Web Analytics**
  A web content analytics application used by IMR using the ASAP programming model that scales to big, heterogeneous data
  www.github.com/project-asap/web-analytics
- **Spark Nested**
  The recursive analytics query module and the hierarchical data decomposition module extensions for the Spark analytics engine.
  www.github.com/project-asap/Spark-Nested

[13] Brasoveanu, A., Sabou, M., Scharl, A., Hubmann-Haidvogel, A. and Fischl, D. (2017). Visualizing Statistical Linked Knowledge for Decision Support, *Semantic Web Journal*, 8(1): 113-137.

- **Spark Tests**
  Set of Apache microbenchmarks to test an alternative scheduling mechanism.
  www.github.com/project-asap/spark-tests
- **Swan Tests**
  Test cases for Swan in clang and intel-cilkplus-runtime
  www.github.com/project-asap/swan_tests
- **Swan Clang**
  Swan clang front-end
  www.github.com/project-asap/swan_clang
- **Swan Runtime**
  Swan runtime - Extension of Intel Cilk Plus runtime with dataflow dependencies between tasks
  www.github.com/project-asap/swan_runtime
- **Swan LLVM**
  Minor update of LLVM including a re-usable routine used in swan_clang
  www.github.com/project-asap/swan_llvm
- **Statistical Tests**
  Tests for the webLyzard Statistical Data API
  github.com/project-asap/statistical-tests
- **extensible Web Retrieval Toolkit (eWRT)**
  Modular open-source Python API to retrieve social data from Web sources, including various helper classes for effective caching and data management.
  www.github.com/project-asap/ewrt
- **infovyz**
  Modular visualization component library based on d3.js. It provides an API optimized for usage in self-updating real-time dashboards. Currently it offers bar charts, line charts and geographic maps.
  www.github.com/project-asap/infovyz
- **ASAP Dashboard**
  Public showcase of the various ASAP visualizations as well as the high-performance dashboard synchronization mechanisms; includes selected datasets to test the on-the-fly drill down operations.
  asap.weblyzard.com

5. Project Consortium

Foundation for Research and Technology – Hellas (FORTH)

FORTH is an internationally acknowledged research center, its Institute of Computer Science (FORTH-ICS) being ranked first among all Greek institutes in its field in two national evaluation rounds. The CARV laboratory of FORTH-ICS has a 25-year history in architecture, hardware, and systems software research, with fundamental contributions in cluster computing, inter-processor communication, scalable systems, I/O subsystems, infrastructure services engineering, parallel and distributed algorithms, and parallel programming languages. FORTH currently coordinates approximately 30 FP7 projects. Within ASAP, FORTH leads the development of the analytics engine as well as the system integration, and coordinates the consortium.

- Role. Project Coordinator. Runtime systems, programming models, scheduling, scalability, execution engines.
- Key Personnel. Polyvios Pratikakis, polyvios@ics.forth.gr

Université de Genève (UNIGE)

UNIGE participates in the project with the Institute of Services Science (ISS), part of the Centre Universitaire d'Informatique (CUI), an inter-faculty center dealing with different aspects of computing. This multidisciplinary aspect is concretized by significant applications, important publications and several projects in areas as far as Humanities and Biotechnology, Wireless Communication and Services Sciences, Virtual Reality and Software Engineering, Multimedia and Computer Vision. The CUI members have participated in and led numerous EU research projects. The role of the University of Geneva in ASAP is to lead the research on the adaptation of data analytics, through the management of user query workflows input to the system.

- Role. Data storage, query planning, adaptation and execution monitoring
- Key Personnel. Verena Kantere, verena.kantere@unige.ch

Institute of Communication and Computer Systems (ICCS)

The Computing Systems Laboratory (CSLab, www.cslab.ece.ntua.gr), part of the Institute of Communication and Computer Systems (ICCS) of NTUA (www.iccs.gr/eng), is one of the largest research laboratories in the Computer Science Department of the School of Electrical and Computer Engineering, National Technical University of Athens. It employs experienced staff in administration, training, consulting and development covering aspects of large-scale high performance computing, and networking and storage technology.

ICCS has large involvement in several national and European research projects and possesses a strong expertise in data management aspects of large scale distributed systems. ICCS undertakes the role of the Technical Coordinator in the project and leads the design and implementation of the central Intelligent Resource Scheduling (IReS) platform that models, decides and schedules analytics jobs.

- Role. Data storage, query scheduling, cost modeling, elasticity
- Key Personnel. Dimitrios Tsoumakos (dtsouma@cslab.ece.ntua.gr), Katerina Doka (katerina@cslab.ece.ntua.gr)

Queen's University Belfast (QUB)

Established in 1845, QUB is the 9th oldest university in the UK and a member of the Russell Group, the Ivy League of the top research-driven universities in the UK. The activities of ASAP are hosted at the High Performance and Distributed computing (HPDC) Research Cluster. Cluster expertise spans all layers of system software (languages, compilers, runtime systems, OS/virtualization framework), workload characterization, parallel foundations of concurrency, and infrastructure support for large-scale, data-intensive applications. The role of QUB in the ASAP project was to lead the definition of the unified programming language that can be used productively by data analysts to express algorithms in data analytics.

- Role. Programming models, runtime systems, execution engines, scheduling, high performance computing
- Key Personnel. Hans Vandierendonck (h.vandierendonck@qub.ac.uk), Dimitrios Nikolopoulos (d.nikolopoulos@qub.ac.uk)

Internet Memory Research (IMR)

IMR is a spinoff of the Internet Memory Foundation dedicated to Web archiving and Web-scale extraction of information for professional use. In addition to its archiving services, IMR develops tools to collect and extract information at a large scale (millions of sites and social networks) for professional Web intelligence applications. While currently crawling dozens of terabytes of data per month, IMR plans to scale to petabytes in the near future. IMR team members have extensive experience in FP6 and FP7 projects. This includes the STREP projects LIWA and LAWA on large scale web analytics for research, and the FET IP Living Knowledge. The ASAP project supports the development of concurrent and continuous Web content analytics, to be integrated in the *Mignify.com* platform.

- Role. Web analytics, data center infrastructure, analytics engineering
- Key Personnel. Philippe Rigaux, philippe.rigaux@internetmemory.net

Wind Tre SpA (WIND)

Since 2011, Wind is part of the VimpelCom Group, the 5th largest mobile TLC operator in the world. In 2016, Wind Telecomunicazioni SpA and H3G SpA executed a merger, changing their corporate name to Wind Tre SpA - following the 50/50 joint venture created by CK Hutchison and VimpelCom. Wind Tre is the leading mobile telecom operator in Italy, with over 31 million mobile customers and 2.7 million in the fixed line. Wind Tre owns and manages a vast amount of data originating from millions of daily streams. It uses these data to optimize services, detect anomalies and improve the network. The ASAP framework supports these tasks and improves Wind Tre's capabilities in dealing with new kinds of innovative applications based on the knowledge arising from big data. Wind Tre harvested its expertise in telecommunications data management and analytics to lead the efforts in WP9 "Applications: Telecommunication Data Analytics", and contributed to the dissemination/exploitation and system requirements tasks of the project.

- Role. Telecommunications analytics, marketing analytics, analytics engineering, data center infrastructure
- Key Personnel. Maria Rita Spada (mariarita.spada@windtre.it), Roberto Bertoldi (roberto.bertoldi@windtre.it)

webLyzard technology (WLT)

WLT provides advanced Web intelligence services to analyze the perceptions of stakeholders (customers, journalists, policy makers, etc.). Real-time content feeds from news and social media contribute to an evolving knowledge repository, accessible via an award-winning media analytics dashboard. The dashboard enables decision makers to identify relevance and sentiment of online content, measure brand reputation, and apply advanced communication success metrics. Innovations from the ASAP research project will enable novel use cases and extend the capabilities of existing applications such as the UNEP Live Web Intelligence[14] platform and the semantic search of the U.S. Climate Resilience Toolkit.[15]

- Role. Information visualization, knowledge extraction, analytics engineering
- Key Personnel. Arno Scharl (scharl@weblyzard.com), Walter Rafelsberger (rafelsberger@weblyzard.com)

[14] www.weblyzard.com/unep-live

[15] www.weblyzard.com/climate-resilience-toolkit

FP7 Project ASAP
Adaptable Scalable Analytics Platform

End of D10.7 Showcase

WP 10 – Exploitation and Dissemination
ASAP Consortium

Nature: Report
Dissemination: Public