

**FP7 Project ASAP**  
Adaptable Scalable Analytics Platform



# D10.4 Yearly Report for Exploitation and Dissemination of Foreground

This report describes the strategy, channels and methods for exploitation and dissemination of project results, with the goal of maximizing the impact of ASAP

## WP 10 - Exploitation and Dissemination

**Nature:** Report

**Dissemination:** Public

**Date:** 28 Feb 2015

### Version History

Version	Date	Author	Comments
0.1	10 Jan 2015	A. Scharl	Initial Version based on D10.3
0.2	25 Jan 2015	A. Scharl	First Revision
0.3	30 Jan 2015	A. Scharl	Second Revision
0.3.x	01-15 Feb 2015	N.N.	Various Contributions by Consortium Partners
1.0	28 Feb 2015	A. Scharl	Final Revision and Layout

### 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.

**Table of Contents**

1. Executive Summary .....	3
2. Knowledge Creation and Dissemination.....	4
2.1 General Dissemination Activities .....	4
2.1.1 ASAP Logo.....	4
2.1.2 Project Web Site .....	4
2.1.3. Social Media Channels .....	6
2.1.4. ASAP Fact Sheet .....	7
2.1.5 Other Promotional Activities .....	8
2.2. Scientific Publications and Related Public Outreach Activities .....	8
2.2.1. Submitted and Planned Papers.....	9
2.2.2. Events and Presentations .....	11
2.3. Media Coverage .....	12
2.4. Research Collaboration .....	12
2.5. Software and Data Resources.....	13
2.6. Industry-Oriented Materials .....	13
3. Exploitation.....	14
3.1. Internet Memory Research .....	14
3.2. WIND Telecomunicazioni .....	15
3.3. webLyzard technology .....	17

# 1. Executive Summary

This report summarizes the **exploitation** and **dissemination** activities in Y1 of the ASAP project, describes future activities planned for Y2 and Y3, and outlines how the project results will be made available to the **European industry**. This document will be updated during the project (D10.4-D10.6).

All partners will continue to play active roles in dissemination, placing special emphasis upon **scientific channels** of publications including journal and conference papers, but also actively engaging in workshops, summer schools and other courses, invited talks, and various online dissemination activities. External **networks** and **advisors** will form a key part of the dissemination strategy towards non-scientific users, as they are providing **outreach** towards **citizens and organizations** in diverse sectors - e.g., press releases, hosting of events, referencing the project in online publications, and presentations aimed at non-scientific users.

The main public interface of ASAP is the project **web site** ([www.asap-fp7.eu](http://www.asap-fp7.eu)), which contains information on the project objectives, partners, R&D activities, and summaries of results as they become available. Furthermore, the scientific results of the project are being submitted to major international **conferences** and **journals**. The ASAP project partners will take part in relevant national and European concertation events.

In order to promote maximum use and dissemination, major technology components from ASAP will be published under an **open source** license and are thus easily exploitable both commercially and for research purposes. In order to ensure support beyond the project's lifetime, the source code of these components will be made available on well-known public platforms such as **GitHub.com** or **SourceForge.org**.

## 2. Knowledge Creation and Dissemination

For maximum impact, ASAP has adopted a **multi-channel dissemination approach**, which will be outlined in the following - this includes setting up online spaces for supporting communities of interest, participating in scientific conferences and exhibitions, as well as publishing at high impact journals and conferences.

These dissemination activities are aimed at creating a widespread **awareness** and **understanding** of the benefits of ASAP, interacting with stakeholders and demonstrating the work to businesses and **potential customers**, sharing **technological achievements** with researchers and practitioners, establishing and maintaining a favorable reputation of the project, and fostering further collaborations and exploitation activities.

The consortium recognizes that a favorable reputation created through dissemination can lead to **competitive advantage** and generate market demand for the products or services being created and exploited. This will support the participating companies who are actively engaged in implementing and delivering commercial software offerings based on ASAP.

The following sections will first describe **general dissemination activities** (e.g., designing a logo, setting up the project Website), and then describe the dissemination plans of individual **project partners**, who will contribute to project dissemination in ways that fit their research domain and technology development.

### 2.1 General Dissemination Activities

#### 2.1.1 ASAP Logo

The following logo below has been designed for ASAP and is used in all dissemination materials; it is provided as a shaded bitmap version (left) as well as a black-and-white vector version (right) for high-quality print reproduction.



#### 2.1.2 Project Web Site

The project has set up and maintains a public **Website** at [www.asap-fp7.eu](http://www.asap-fp7.eu), allowing distributed editing among consortium members. Following an initial static HTML version, a dynamic site based on the WordPress content management system with a

professional theme has been activated in August 2014. WordPress supports a more interactive development process and the independent editing of document by consortium partners.

This Website functions as a project dissemination and documentation tool and single access point that references other ASAP **applications** and **services**. In addition to documenting the scientific output of the project, examples and descriptions will outline how the information services can be used by third parties, independently or in conjunction with other Web applications.



## A Scalable Analytics Platform (ASAP)

The ASAP FP7 research project develops 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 (and store) is suitable for all types of data. Complex analytical tasks over multi-engine environments therefore require integrated profiling, modeling, planning and scheduling functions. Addressing these challenges, ASAP pursues four main goals:

1. A generic **task-parallel programming model** in conjunction with a runtime system for execution in the cloud. The runtime will incorporate and advance state-of-the-features including:
  - irregular general-purpose computations,
  - resource elasticity,
  - synchronization, data-transfer, locality and scheduling abstraction,
  - ability to handle large sets of irregularly distributed data, and
  - fault-tolerance.
2. 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.
3. A **unique adaptation methodology** that will enable analytics experts to amend submitted tasks in later processing stages.
4. A **real-time visualization engine** to show the results of the initiated tasks and queries in an intuitive manner - building on the dashboard of the [Media Watch on Climate Change](#) and the faceted search developed for the [Climate Resilience Toolkit](#).

- News
- Overview
- Use Cases
- Publications
- Presentations
- Deliverables
- Consortium

### PROJECT FACT SHEET

<p><b>NAVIGATION</b></p> <ul style="list-style-type: none"> <li>News</li> <li>Overview</li> <li>Use Cases</li> <li>Publications</li> <li>Presentations</li> <li>Deliverables</li> <li>Consortium</li> </ul>	<p><b>PROJECT SUMMARY</b></p> <p>The ASAP FP7 Research Project develops a dynamic open-source execution framework for scalable data analytics, and applies this framework to two use cases focusing on Web content and telecommunications data. ASAP assumes that no single execution model is suitable for all types of tasks, and no single data model (and store) is suitable for all types of data.</p>	<p><b>ACKNOWLEDGEMENT</b></p> <p>This project receives funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number 619706.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

[www.asap-fp7.eu](http://www.asap-fp7.eu) will be updated continuously during the project, as progress is made and new outputs become available. The content of the Website includes:

- Project overview and objectives;
- Project partner information;
- Downloadable project outputs (publications, deliverables, etc.);
- News and links to dissemination events with deadlines, related projects and research activities, etc.
- Links to related projects and research activities;
- Demos that have been added under scientific results section.

### 2.1.3. Social Media Channels

Posting project-related information to social media platforms will extend outreach and increase the exposure of project results. The ASAP project has now established a Twitter account (**@ASAP\_EU**), where all project partners will disseminate their publications, events organised and participated in, etc.

Project members also individually disseminate ASAP-specific results in additional social media channels such as LinkedIn and Facebook. Either the account name **@ASAP\_EU** or the hash tag **#asap-fp7** will be used to tag posts and material published via these channels.



## 2.1.4. ASAP Fact Sheet



## ASAP Research Project - A Scalable Analytics Platform

[www.asap-fp.eu](http://www.asap-fp.eu) • [@ASAP\\_EU](https://twitter.com/ASAP_EU)

The **ASAP FP7 Project** develops 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 (and store) is suitable for all types of data.

Complex analytical tasks over multi-engine environments therefore require integrated profiling, modeling, planning and scheduling functions. Addressing these challenges, ASAP pursues four main goals:

- A general-purpose task-parallel programming model in conjunction with a runtime system for execution in the cloud. The runtime will incorporate and advance state-of-the-art features including (i) irregular general-purpose computations, (ii) resource elasticity, (iii) synchronization, data-transfer, locality and scheduling abstraction, (iv) the ability to handle large sets of irregularly distributed data, and (v) 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.
- A unique adaptation methodology that will enable analytics experts to amend submitted tasks in later processing stages.
- A real-time visualization engine to show the results of the initiated tasks and queries in an intuitive manner – building on the webLyzard Web intelligence dashboard, which can be tested by accessing the *Media Watch on Climate Change* ([www.ecoresearch.net/climate](http://www.ecoresearch.net/climate)).

### Use Cases and Applications

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

#### USE CASE 1 – WEB CONTENT ANALYTICS

The services of *Internet Memory Research* as part of the Mignify platform ([www.mignify.com](http://www.mignify.com)) provide access to a very large Web content collection – cleaned, annotated and indexed in a distributed infrastructure mainly based on Hadoop components. ASAP will extend and enrich the public workflow interface supplied by Mignify, referred to as pipes (queries associated with a set of intelligent agents to extract or transform large-scale Web data). ASAP will extend the pipe specification with iteration and fixpoint primitives to support three scenarios:

- **Public Interface** to let customers specify and execute pipes for Web content. All pipes have to run concurrently within a single distributed infrastructure. It is essential to schedule and coordinate the execution of pipes to obtain a reliable estimate on pipe execution time, and to report expected response times.

- **Infrastructure** to store data and run pipes. Many distributed applications share the resources of the IMF infrastructure, which will require a scheduling module to allocate resources for pipe execution based on the services' constraints.
- **Stream Processing.** In many scenarios - e.g., the extraction of indicators from social media sources, the pipe should run almost continuously on incoming content. ASAP will study how such a continuous subscription mechanism can be implemented in the context of a large number of concurrent workflows.

#### USE CASE 2 – TELECOMMUNICATIONS DATA

*Call Detail Records* (CDR) data is a good proxy to understand human mobility. The sheer volume of this data poses new challenges when extracting and visualizing specific indicators. ASAP will investigate the following applications:

- **Event Detection** to analyze the different features of an event, including its spatio-temporal characteristics, social aspects, and statistical properties. By controlling input parameters such as the time interval, the spatial area and additional CRM attributes, analysts gain a detailed understanding of evolving events.
- **Ridesharing** provides functions for mobility managers and individual drivers alike, for example 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 specific 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 and with the option to provide spatio-temporal constraints as input.

### Project Partners

- Foundation for Research and Technology – Hellas  
[www.forth.gr](http://www.forth.gr)
- Université de Genève  
[www.unige.ch](http://www.unige.ch)
- Institute of Communication and Computer Systems  
[www.iccs.gr/eng](http://www.iccs.gr/eng)
- Queen's University Belfast  
[www.qub.ac.uk](http://www.qub.ac.uk)
- Internet Memory Research  
[www.mignify.com](http://www.mignify.com)
- WIND Telecomunicazioni  
[www.wind.it](http://www.wind.it)
- webLyzard technology  
[www.weblyzard.com](http://www.weblyzard.com)



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 619706.

### 2.1.5 Other Promotional Activities

- webLyzard organized the printing of 4,000 business cards using the corporate identity established through the WordPress theme of the ASAP Web site. The cards are available to all project partners to increase project visibility at conferences and workshop. They serve as a cost effective, environmentally sustainable and often more accepted alternative to regular printing material.



- FORTH organized the printing of T-Shirts with the ASAP logo as an additional promotional item to increase project visibility, for example when attending conferences or showcasing the project at various events.



## 2.2. Scientific Publications and Related Public Outreach Activities

To be documented on the ASAP Website, we will publish a growing body of **scientific papers** and **presentations** at **technical and scientific conferences**. The specific outlets will be selected based on their relation to the topics addressed by ASAP, and on their impact. The publications are aimed at sharing the results of ASAP with



the European and international scientific community, to encourage their incorporation into the work of other scientists and technologists. The project will also provide online software demonstrations, available from the ASAP Website.

With respect to IPR and promoting take-up of project results, we endeavour to make all **scientific results** of this project available to the research community. In addition, associated tools will be released as open-source software to facilitate take-up.

In order to ensure the widest possible audience, we will use **press releases** and **training activities** aimed at the academic and industrial sectors (e.g., delivery of tutorials), industry-oriented dedicated workshops, etc. In May 2014, WLT has sent out a newsletter (n=642)<sup>1</sup> and published a corresponding document on its Website.<sup>2</sup> The ASAP project has also been mentioned in the June 2014 issue of "Swiss CIO", a publication by the Swiss Computerworld (see Section 2.2.1).

### 2.2.1. Submitted and Planned Papers

The ASAP consortium aims to actively produce and disseminate research results, targeting top-tier **international conferences** and **journals** such as VLDB, SIGMOD, ICDE, CIKM, BigData, Cloud, IEEE TKDE and ACM DKE.

#### *Accepted or Submitted Refereed Publications*

#### **2015**

- Doka, K., Papailiou, N., Tsoumakos, D., Mantas, C. and Koziris, N. (2015). IReS: Intelligent, Multi-Engine Resource Scheduler for Big Data Analytics Workflows. To appear in Proceedings of the 2015 ACM SIGMOD/PODS International Conference on Management of Data (SIGMOD'15, Demo track), Melbourne, Victoria, Australia.
- Mytilinis, I., Tsoumakos, D., Kantere, V., Nanos, A. and Koziris, N. (2015). "I/O Performance Modeling for Big Data Applications over Cloud Infrastructures". IEEE International Conference on Cloud Engineering (IEEE IC2E 2015).
- Papailiou, N., Tsoumakos, D., Karras, P. and Koziris, N. (2015). Graph-Aware, Workload-Adaptive SPARQL Query Caching. To appear in Proceedings of the 2015 ACM SIGMOD/PODS International Conference on Management of Data (SIGMOD'15), Melbourne, Victoria, Australia.

---

<sup>1</sup> [www.eepurl.com/TsvIn](http://www.eepurl.com/TsvIn)

<sup>2</sup> [www.weblyzard.com/scalability](http://www.weblyzard.com/scalability)

- Papakonstantinou, Zakkak, Pratikakis (2015). A Parallel Task Execution Engine (Submitted for review).
- Vandierendonck, H. (2015). Compiler and Runtime Support for Hybrid Static/Dynamic Scheduling. Workshop on Compilers for Parallel Computing (CPC-15).
- Vandierendonck, H. (2015). Map-Reduce on Shared-Memory Systems: Has Specialization Paid Off? Submitted to: International Conference on Supercomputing. 10 pages.

## 2014

- Pavlos Katsogridakis, Polyvios Pratikakis (2014). Hama and Hadoop: A Performance Comparison. In Summer School on Advanced Computer Architecture and Compilation for High-Performance and Embedded Systems (ACACES).
- Kantere, V. (2014). "A Holistic Framework for Big Scientific Data Management", International Congress on Big Data (IEEE BigData 2014).
- Kantere, V. (2014). "Mapping Construction Compliant with Schema Semantics", International Conference on Database and Expert Systems Applications (DEXA 2014).

### *Trade Publications*

- Scharl, A. and Weichselbraun, A. (2014). "Die Meinungs-Erforscher", Swiss CIO Magazine, Computerworld, 20 June 2014, 32-35.

### *Future Publication Plans*

**ICCS** has already published 2 research papers and a demo in top-tier international conferences and intends to further disseminate project results via the submission of research papers and the publication of articles in specialized scientific conferences and journals. The related research areas include Big Data, Distributed Computing, Cloud Computing, Data Management, Performance Modeling, etc. In the course of the project the number of submissions will increase, as soon as the first tangible results are available.

**IMR** mostly participates in workshops and seminars where the R&D team is invited to present technical achievements, performance evaluations, use cases and industrial assessments of open source tools for Big Data management (see, e.g., our presentation at Big'14, a workshop collocated with WWW'14). ASAP will typically supply material and results that could be presented during such events.

**UniGe** has started disseminating the initial research results to international scientific conferences. Two papers have been accepted for publication: (i) "A Holistic Framework for Big Scientific Data Management" has been accepted to the International Congress on Big Data (IEEE BigData), 2014; (ii) "Mapping Construction Compliant with Schema Semantics" has been accepted to the International Conference on Database and Expert Systems Applications (DEXA), 2014.

**webLyzard.** Preliminary results of scalability improvements in the WLT knowledge extraction processing pipeline are currently under review. WLT plans to submit further results in the second quarter of 2015, particularly in regards to the progress made in WP6 (see D6.1 on the visualization of linked statistical data). In addition to the examples listed at the beginning of Section 2.2.1, we will also target other ACM and IEEE outlets, as well as dedicated journal calls focusing on information management, visual analytics, dashboard design, and human-computer interaction.

**FORTH** will disseminate the scientific results of the project via research paper conference publications, journal publications, workshop presentations, and poster presentations in related scientific venues. Moreover, FORTH will organize and participate in sessions in related events such as network-of-excellence meetings. Related areas include High-Performance Computing, Big Data, Distributed Systems, Parallel Programming and Programming Languages. FORTH is currently organizing a thematic session in the HiPEAC Computing Systems Week event in October 2014.

### 2.2.2. Events and Presentations

#### 2015

- Jan 2015. Kantere, CLUSIS Conference 2015.
- Jan 2015. Tsoumakos, H2020 ICT-16 Big Data Research, Information and Networking Event.
- Jan 2015. Vandierendonck, Workshop on Compilers for Parallel Computing.

#### 2014

- Nov 2014. Kantere, ANR Cloud-based Organizational Design conference (CBOD 2014).
- Aug 2014. Vandierendonck, Summer School on Advances in Programming Languages (AiPL 2014).
- Jun 2014. Pratikakis, 10th Int'l Summer School on Advanced Computer Architecture and Compilation for High-Performance and Embedded Systems (ACACES 2014).

### 2.3. Media Coverage

- Nov 2014. Weißes Haus setzt bei neuer Klima-Initiative auf österreichische Web Intelligence | OekoNews AT.
- June 2014. Die Meinungs-Erforscher | Computerworld CH.

### 2.4. Research Collaboration

**ICCS** will exploit results from the recently concluded FP7 Project **ARCOMEM** ([www.arcomem.eu](http://www.arcomem.eu)). Within ARCOMEM, ICCS team members have developed the ARCOMEM database, a component responsible for storing huge amounts of crawled raw data as well as metadata of annotated web archives. The invaluable experience gained in various data stores such as HDFS, HBase, H3RDF (a distributed rdf store developed for the project) will be exploited in accurately modelling such engines.

The **CELAR** project ([www.celarccloud.eu](http://www.celarccloud.eu)) currently works into elasticity aspects of applications deployed over cloud infrastructures. **ICCS** will exploit its joint involvement in the two projects to evolve research and tools relevant to profiling, modeling and elasticity decisions customized to Big Data and analytics engines.

**IMR** has been closely working with TU Berlin, during several years, on the applicability of Stratosphere, a distributed data processing system, to web data management. Since April 14<sup>th</sup>, Stratosphere is an official Apache project, named Apache Flink. We are keen to develop our experience Flink in the context of ASAP.

The **CACTOS Project** ([www.cactosp7.eu](http://www.cactosp7.eu)) develops the data analytics component of CactoScale, a set of tools and methods to acquire and analyse application behaviour and infrastructure performance data. CactoScale conducts detailed platform and service workload characterisation by mining console and system performance logs. The mined information informs optimisation strategies for VM placement and migration within and across datacentres. The analytic tool of CactoScale can directly feed into ASAP to support the mapping of application components to data centres and services in a way that exploits their topological properties and is aware of application execution patterns and resource demand.

**webLyzard**. ASAP benefits from the results of related national research projects such as **DIVINE** ([www.weblyzard.com/divine](http://www.weblyzard.com/divine)). We will also pursue active research cooperations, identify synergies to maximize impact, and exchange results with other ongoing projects such as **uComp** ([www.ucomp.eu](http://www.ucomp.eu)), which develops a human computation framework to blend knowledge extraction and crowdsourcing approaches, **DecarboNet** ([www.decarbonet.eu](http://www.decarbonet.eu)), which builds a collective awareness platform for carbon footprint reductions, and **PHEME** ([www.pHEME.eu](http://www.pHEME.eu)), which tackles *veracity* as an emerging challenge of big data research - in addition to Gartner's 3Vs of big data: *volume*, *velocity*, and *variety*.

## 2.5. Software and Data Resources

Software prototypes will be made publicly available for download from the project Website or public repositories such as GitHub.com and SourceForge.org. The list of currently available components includes:

- The **Swan Task Dataflow Scheduler** and its (preliminary) distributed variant [www.github.com/hvdieren/swan](http://www.github.com/hvdieren/swan) [www.github.com/Richard-Neill/distributedswan](http://www.github.com/Richard-Neill/distributedswan)
- **easy Web Retrieval Toolkit (eWRT)**  
[www.github.com/weblyzard/ewrt](http://www.github.com/weblyzard/ewrt)  
Modular open-source Python API to retrieve social data from Web sources, including various helper classes for effective caching and data management.
- **Media Watch on Climate Change (MWCC)**  
[www.ecoresearch.net/climate](http://www.ecoresearch.net/climate)  
Publicly accessible showcase of the webLyzard dashboard with data export capabilities (extended through T6.1), focusing on climate science and related environmental issues.
- **Profiler for Modeling Application Performance over Virtualized Resources (PANIC)**  
<https://github.com/giagiannis/panic>
- IMR's policy does not include direct participation to open source projects, although we indirectly contribute to the development and assessment of open source software such as **Apache Flink**.

## 2.6. Industry-Oriented Materials

As a result of ongoing industry-oriented activities, exploitable results of ASAP will be made public through the ASAP Website as software packages, presentations and dedicated web documents. The industrial partners within the ASAP consortium will seek synergies and emphasize commonalities in industry-oriented presentations. We will pursue a more differentiated approach vis-à-vis specific sectors, depending on which organization is most skilled to address this sector.

The presentations are aimed at helping convince early adopters about the benefits and opportunities that ASAP can offer. Therefore, the presentations will include a comparison of existing solutions and collect specific business needs from industry.

## 3. Exploitation

Due to the generic nature of the project results, especially the ASAP unified programming model and distributed computing engine, exploitation activities will go beyond a specific industry and beyond the defined uses cases. Tailored exploitation actions will be targeted at **companies** already collaborating with partners of the ASAP consortium (in various domains, not exclusively those covered by the two use case), **working groups** and **standardization** bodies, and other **stakeholders** with an interest in big data technologies such as policy makers and NGOs.

### 3.1. Internet Memory Research

IMR develops a Web data analytics platform called **Mignify**, which gives customers access to a large collection of contents crawled from the Web, periodically refreshed, cleaned, and stored/indexed in a distributed repository based on Hadoop. IMR publishes **user content** extracted from active sources (blogs, forums, etc).

Customers can get these contents through **pipes**, which combine queries to retrieve data of interest, and text mining / data mining workflows applied to the query results. The use case proposed to our ASAP partners focuses on improving and extending the current pipe functionalities along several directions.

One of these directions is the **functional extension** of pipes. At the beginning of the project, workflows are essentially limited to a linear application of text mining methods (called *agents*). Modern execution engines support advanced data processing operators for distributed evaluation of iterative algorithms, joins, grouping, etc. This will make possible the execution of data mining and machine learning methods as part of the analysis of web content, and will constitute a highly valuable tool proposed to our customers.

The current model of Mignify assigns an economic “value” to each mining agent, based on the estimated business value supplied to the end user. Our **entity disambiguation** agent for instance is more valuable than the language identification agent, and therefore proposed at a higher price. Bringing agents for clustering, classification, time series analysis, topic modeling, will considerably improve the attractiveness of Mignify services.

Several engines will be investigated during the project. We expect that the collaboration with our expert partners in distributed systems and data processing issues will help us to identify the pros and cons of each solution and to accurately evaluate the circumstances in which a specific engine is a superior choice, with respect to dimensions like **performance**, **robustness**, **ease of deployment**, **cost of implementation** and **maintenance**, etc. Our goal is to publish advanced mining services by the end of the project, thereby improving the attractiveness and visibility of Mignify.

Another direction of R&D that will directly contribute to our platform is related to system aspects. Currently, all our services are hosted in the Mignify **data center**. This constitutes a limited set of resources which should be exploited optimally. Two lines of improvements have been identified here.

- The first one pertains to the **concurrent execution** of pipes in our platform. In the moment, we are restricted to the simple model of one pipe execution at a time. We want to be able to run several pipes concurrently, and to identify indicators (especially memory and CPU resources requirements of a pipe) to design a scheduler that will properly use the available resources. This will greatly improve our return on investment, and quality of our services regarding their availability and timely response to customers' requests.
- Second, being able to **estimate resource consumptions** of a pipe execution might result in several scenarios, with varying costs, proposed to our customers, ranging from low priority / low cost execution, to real time / high value execution. Offering such flexibility, combined with access to cleaned and up-to-date web content and a rich catalog of operators, would highly leverage the current potential of Mignify. ASAP can help us design and test such a cost-based approach.

### 3.2. WIND Telecomunicazioni

WIND contributes to communicating the project results by via internal corporate channels as well as through the Web-based channels of the Vimpelcom group. Furthermore, WIND exploitation activities are dealing with the impact of the ASAP project in the area of the **privacy-aware mobility mining** to improve the portfolio of services, taking into account all the acquired knowledge in a big data context.

ASAP will enable WIND to quickly develop telecommunications analytics applications similar to the application described in WP9, at a much lesser cost. As telecommunications organizations have access to big data, offering a lower-cost solution than current alternatives for developing analytics applications using that data, gives a competitive advantage, reducing the time and cost of development (for analytics queries and applications, automating the scheduling process) and the overhead costs by increasing scalability and thus making better use of resources.

WIND has proven technology capabilities in the areas used in the various network segments, ranging from mobile and landline access to the backbone transmission network with huge clientele in the order of tens of millions of users. As part of its strategic evolution, Wind is actively pursuing on-going activities on innovative wireless technologies which enable offering new services in different conditions and environments. All of those activities will be highly enhanced through the use of a powerful, **unified analytics platform**, enabling easier than ever large scale data analysis aid-

ing at trend discovery and smart decision making. WIND plans on evaluating the effect of the **tourist trends application** in order to further invest on similar analytics applications that cover more business sources and also fuse them with the 'wisdom of the crowds' (i.e., social data). Here are some details on how ASAP with its specific orientation can offer innovative implications to industrial performance.

The objective of the project is actually to develop an innovative forecasting model of **tourist flows** linked to events and places, combining the analysis of traffic data coming from the telephone network with information available on social networks. The whole process follows a privacy-preserving approach that guarantees compliance to national privacy regulations.

Previous experience with phone data has revealed their ability to highlight the fact that a particular event is taking place, providing estimates on the number of persons in a given area and showing correlations with different areas. On the other hand, data from **social networks** is able to provide assessments of the interest in a particular event (past, present or planned) and other semantic information derived from the text in the "posts" published by the participants.

Once developed this model has several potential applications and can be useful in various contexts such as: (i) **transport** - the (re-)construction and optimization of routes, (ii) **urban planning** - promoting the area with customized and "smart" solutions, and (iii) **social networks** - e.g. analysis of tourism in the area by studying the behavior and movements of people in that particular area

From the industrial point of view the application that will be developed, even if at prototype level, will bring to light how the strategic use of **big data** and modern analysis techniques will be useful to offer a service of **enriched analysis** able to quantify the rate of success of an event and improve the effectiveness of services released in the context of the selected event.

In addition, the method to discover the value of the data during the pilot phase can help in the creation of a new culture of the quality and use of the data and to establish an innovative approach to the design and development of applications. **Transportation companies** can use the data coming from planes, trucks and carriers to optimize routes, maximize truckloads and streamline dispatch activities. A distributor can monitor which food and beverages sell well in various locations. By making the right **inventory choices** the merchandising chain can be made more profitable.

Finally, from the point of view of the company, the ASAP activity allows you to test your ability to work with a multi-disciplinary team to create an organized critical mass capable of handling **new services** during the phases of design, deployment and management of the solutions offered by placing the focus on learning by doing.



### 3.3. webLyzard technology

In terms of visibility and engagement in the relevant research communities, and to attract overseas clients (a crucial factor for an SME focusing on large-scale applications of semantic technologies), the communication activities of webLyzard target **government agencies** and **research centers** in Europe and the United States, and large **business-to-consumer brands**. Such brands are a highly valuable asset of organizations. Therefore they are among the primary exploitation targets, and essential for the continued growth of webLyzard and its long-term commercial success.

Consumers who discuss brands on **social media** not only respond to brand communication, but also play a pivotal role in shaping a brand – e.g. when repeating or commenting on a story via Twitter or Facebook. A deep understanding of this process helps to increase brand performance. Given the volume and complexity of the underlying dataset, **visual methods** such as those developed in WP6 of ASAP are the best way to convey such an understanding. Once embedded into the webLyzard **dashboard**, the visualization components support ad hoc data exploration and show updates in real time - since users can define topics on the fly, the comparison is not restricted to brands, but can also include other entities such as products, persons, or organizations.

Word-of-mouth and leveraging close ties with large federal agencies such as the *National Oceanic and Atmospheric Administration* (NOAA) will also be beneficial for disseminating ASAP results to a wide international audience. Improved **scalability** through the project's distributed computing engine will help to increase the knowledge base of webLyzard, attracting new clients and representing an important **competitive advantage** - particularly in conjunction with the new visualization components developed in WP6.

webLyzard strives to combine a thorough **evaluation** of the developed components with an **early exploitation** of developed technology in real-world applications. The *U.S. Climate Resilience Toolkit* serves as a good example of this strategy. The toolkit was developed in response to **President Obama's Climate Action Plan** and Executive Order to help U.S. communities, businesses, and state, local, and tribal governments prepare for, and build resilience to, climate-related changes and extreme events. The webLyzard platform powers the Toolkit's search function, designed to help visitors quickly locate the most relevant content across all the U.S. federal government's web sites. The **bar chart** developed in T6.1 of WP6 has been integrated into the analytics view of this application. This view enables communication experts at the National Oceanic and Atmospheric Administration (NOAA) to monitor the stream of new online publications and continuously improve the Toolkit's **knowledge repository**.

Visualization services are a core value proposition of webLyzard. Early showcases such as the U.S. Climate Resilience Toolkit demonstrate technological leadership and can be used to increase international visibility, and in turn attract additional clients who license the developed technologies. Bolstered by the successful launch of the Toolkit by Vice President Joe Biden, for example, webLyzard has been nominated in the category "Spectacular" of the **US-A-Biz Award 2015**, which is organized by the Austrian Trade Commission Los Angeles to highlight outstanding achievements of Austrian companies in US markets. The media interest created through this award has been used not only to promote webLyzard products and services, but also to point towards the **ASAP project** as a **driver of innovation** behind some of the underlying core technologies.<sup>3</sup> Similar activities to promote ASAP in conjunction with the webLyzard platform are planned for the second year as well.

---

<sup>3</sup> [www.eepurl.com/bbOmKT](http://www.eepurl.com/bbOmKT)