FP7 Project ASAP
Adaptable Scalable Analytics Platform

ASAP D10.5
Yearly Report for Exploitation and Dissemination of Foreground

WP 10 - Exploitation and Dissemination

Nature: Report
Dissemination: Public

Version History

Version	Date	Author	Comments
0.1	07 Dec 2015	A. Scharl	Initial Version based on D10.4
0.2	23 Jan 2016	A. Scharl	Major Revision
0.2x	Feb 2016	All Partners	Various Content Updates
0.3	26 Feb 2016	W. Rafelsberger	Major Revision
1.0	28 Feb 2016	A. Scharl	Major Revision and Formatting
1.1	20 Apr 2016	A. Scharl	Open Source, Impact Metrics, 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 report summarizes the **exploitation** and **dissemination** activities in the first two years of the ASAP project, describes future activities planned for the third year, and outlines how the project results will be made available to the **European industry**. This document is continuously being updated throughout the project (D10.4-D10.6).

All partners play active roles in dissemination, placing special emphasis on **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** 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 **Website**,[1] which contains information on the project objectives, partners and research and development activities. Supported by the ASAP **Twitter channel**,[2] it also reports new results as they become available. The scientific results of the project are being submitted to major international **conferences** and **journals**, and presented in relevant national and European **concertation events**.

In order to promote maximum use and dissemination, major technology components from ASAP are published under an **Apache open source** license[3] 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 has been made available on **GitHub.com**.[4]

[1] www.asap-fp7.eu
[2] www.twitter.com/asap_eu
[3] www.apache.org/licenses
[4] www.github.com/project-asap

Table of Contents

1. 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 favourable reputation of the project, and fostering further collaborations and exploitation activities.

The consortium recognises that a favourable 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., the use of the logo, the Website and other activities to inform about and promote the project), and then describe the dissemination plans of individual **project partners**, who contribute to project dissemination in ways that fit their research domain and technology development.

1.1 General Dissemination Activities

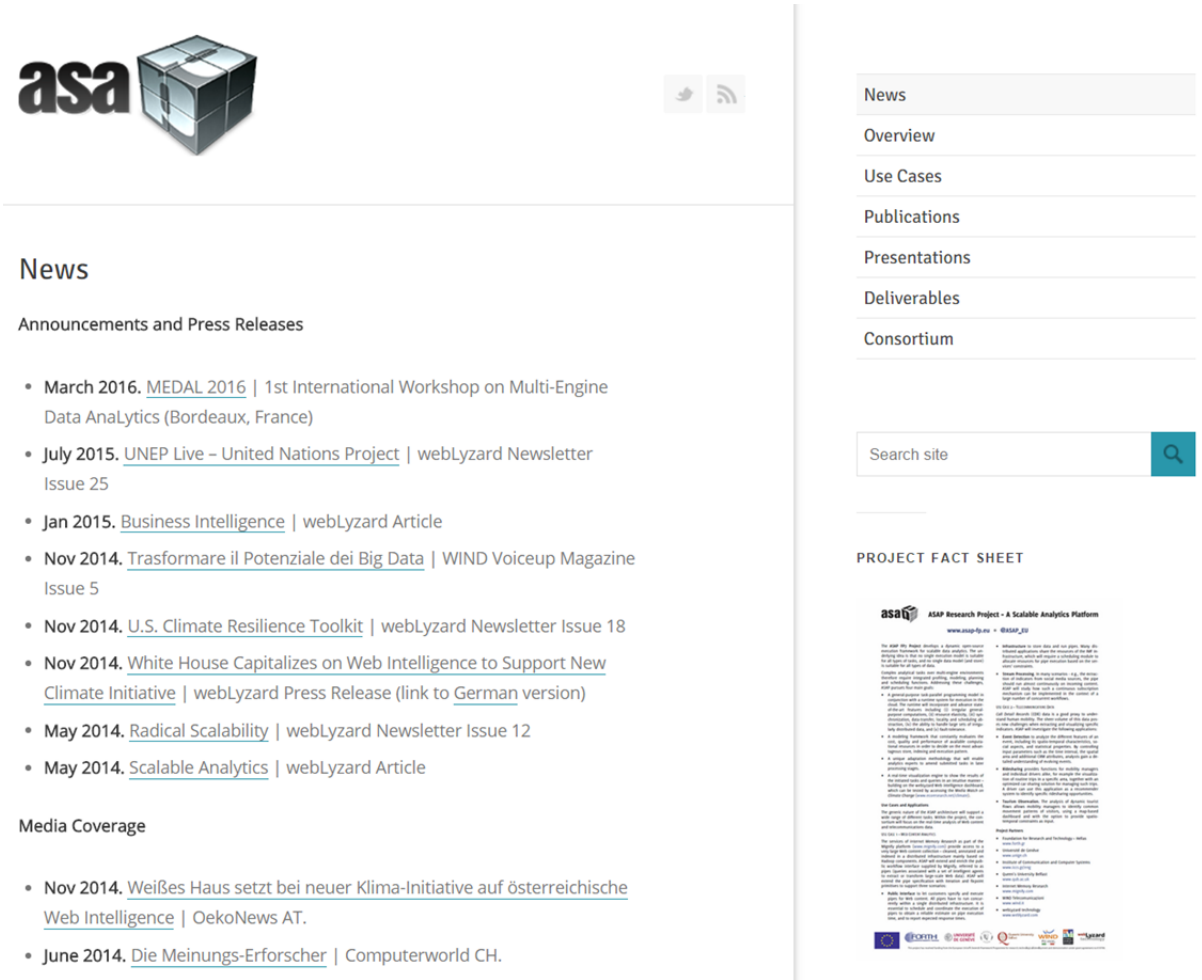
1.1.1 ASAP Logo

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.

1.1.2 Project Website

The project has set up and maintains a public **Website** at **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. An analysis of the log files shows that the Website has attracted more than **4,100 users** in about 4,500 sessions over the last twelve months (= new visitors made up 92% of the total traffic), resulting in about **5,850 page views** as of February 2016. In terms of the users' **geographic origin**, the leading cities were Athens (1.85%), Vienna (1.7%), New York (1.5%) and Belfast (1.0%).

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 outline how the information services can be used by third parties, independently or in conjunction with other Web applications.

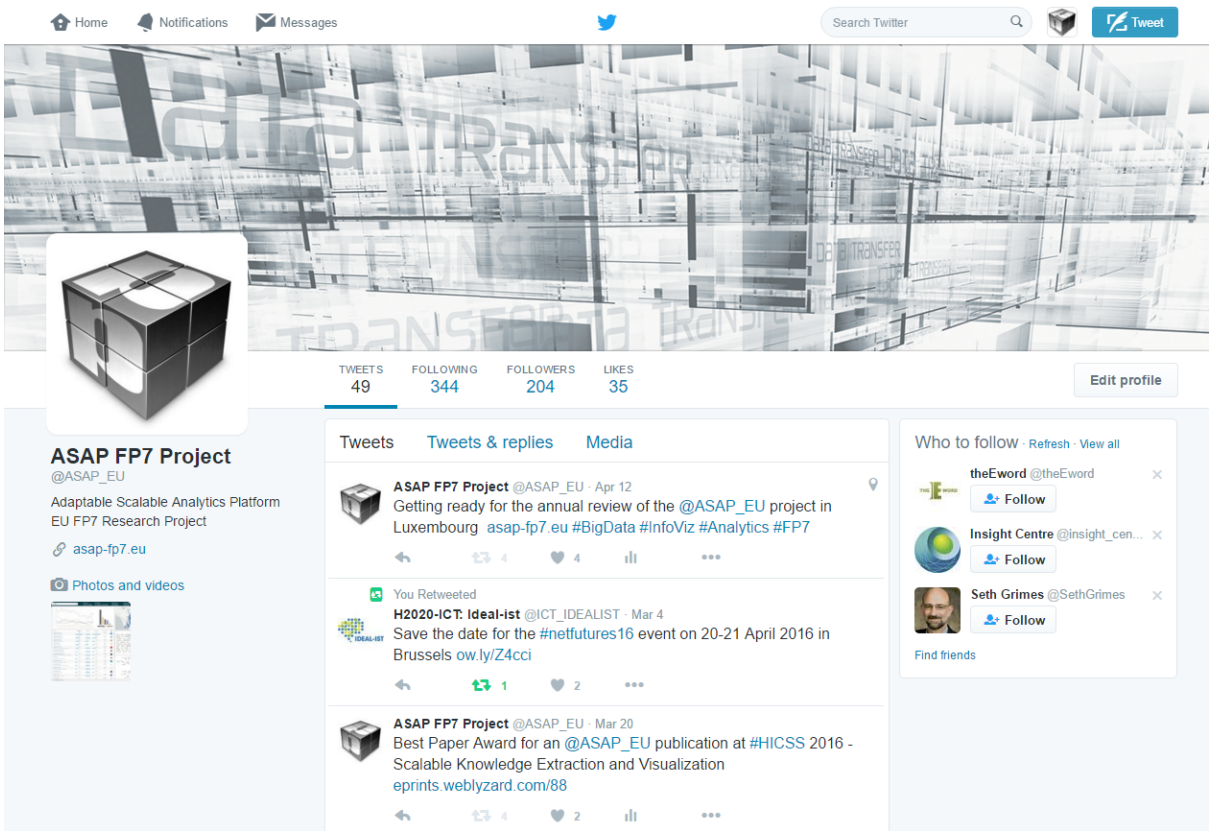


The Website is updated regularly, as progress is made and new outputs become available. The content of the Website includes:

- 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.

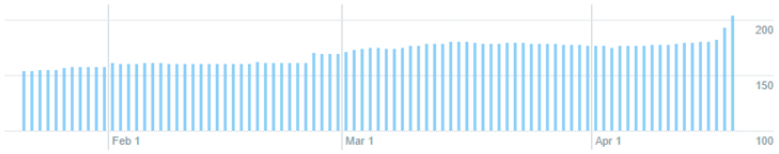
1.1.3 Social Media Channels

Posting project-related information to social media platforms extends outreach and increases the exposure of project results. The ASAP project has established a Twitter account (**@ASAP_EU**), giving the project partners a channel to disseminate their publications, events organised and participated in, etc. As of February 2016, the Twitter account has 170 followers. Project members also 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** is used to tag posts and material published via these channels.



1.1.4 Impact Metrics

The following diagrams reflect the impact of the project's social media activities via the @ASAP_EU account, including follower demographics as well as the number of impressions and engagements of the Top Tweets in the first two quarters of 2016.



Your current follower audience size is **204**
That's 50 more than the same time 90 days ago. You've gained around 1 new followers per day

Interests

Your followers · All Twitter Users

Interest name	Comparison	Difference
Technology	67% / 24%	43% more
Business and news	67% / 43%	24% more
Tech news	66% / 24%	42% more
Business news and general info	55% / 27%	28% more

Gender

71% 29%

Country

Country name	% of audience
Austria	22%
United Kingdom	17%
Greece	12%
Germany	10%

Tweets | Top Tweets | Tweets and replies | Promoted

	Impressions	Engagements	Engagement rate
ASAP FP7 Project @ASAP_EU · Mar 20 Best Paper Award for an @ASAP_EU publication at #HICSS 2016 - Scalable Knowledge Extraction and Visualization eprints.weblyzard.com/88 View Tweet activity	751	28	3.7% Promote
ASAP FP7 Project @ASAP_EU · Apr 12 Getting ready for the annual review of the @ASAP_EU project in Luxembourg asap-fp7.eu #BigData #InfoViz #Analytics #FP7 View Tweet activity	484	10	2.1% Promote
ASAP FP7 Project @ASAP_EU · Apr 17 The @ASAP_EU visual analytics dashboard (query: #Rome); showcases at weblyzard.com/showcases #InfoViz @webLyzard pic.twitter.com/pulqiFoMll View Tweet activity	476	40	8.4% Promote
ASAP FP7 Project @ASAP_EU · Apr 14 Want to recursively spawn new map-reductions from within mapper or reducer tasks in Spark? Use Spark-nesting, part of Project ASAP. View Tweet activity	400	12	3.0% Promote
ASAP FP7 Project @ASAP_EU · Feb 26 Small Big Data Value Association Summit, The Hague 2-3 March 2016 bdva.eu/?q=small-bdva-… @BDVA_PPP View Tweet activity	394	2	0.5% Promote
ASAP FP7 Project @ASAP_EU · Apr 14 ASAP, A scalable analytics platform, now on github. github.com/project-asap View Tweet activity	352	14	4.0% Promote

Engagements
Showing 80 days with daily frequency

ENGAGEMENT RATE
0.8%
Apr 20 4.0% engagement rate

LINK CLICKS
38
Apr 20 1 link click

On average, you earned **0 link clicks** per day

RETWEETS
23
Apr 20 0 Retweets

On average, you earned **0 Retweets** per day

LIKES
28
Apr 20 0 likes

On average, you earned **0 likes** per day

1.2. Scientific Publications and Related Public Outreach Activities

As documented on the ASAP Website, we are publishing a growing body of **scientific papers** and **presentations** at **technical and scientific conferences**. Specific outlets are 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 also provides online software demonstrations, available from the ASAP Website.

With respect to intellectual property rights and promoting take-up of project results, the consortium partners make all **scientific results** of this project available to the research community. In addition, associated tools are released in the form of open source software via **GitHub.com** to facilitate take-up.

In order to ensure the widest possible audience, ASAP uses **press releases** and **training activities** aimed at the academic and industrial sectors (e.g., delivery of tutorials), industry-oriented dedicated workshops, etc. WLT sent out **newsletters** In May 2014 (n=642)[5] and August 2015 (n=753),[6] and published corresponding blog entries on its Website.[7,8] 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). Furthermore, in February 2016 FORTH has sent out a press release resulting in local TV coverage (see Section 2.3).

1.2.1. Publications

The ASAP consortium actively produces and disseminate research results, targeting top-tier **international conferences** and **journals** such as VLDB, SIGMOD, ICDE, CIKM, BigData, Cloud, IEEE TKDE and ACM DKE. Up until now, the project has resulted in a total of **27 refereed publications**:

- 4 journal papers,
- 18 conference papers,
- 3 workshop papers,
- 1 summer school paper, and
- 1 conference demonstration.

The **annual number** of publications has thus **doubled** from 9 to 18 when comparing the first and the second year of the project.

[5] www.eepurl.com/tsvln
[6] www.eepurl.com/bxfe3j
[7] www.weblyzard.com/scalability
[8] www.weblyzard.com/geographic-map

Accepted or Submitted Refereed Publications

2016

- Brasoveanu, A.M.P., Sabou, M., Scharl, A., Hubmann-Haidvogel, A. and Fischl, D. (2015). “Visualizing Statistical Linked Knowledge Sources for Decision Support”, Semantic Web Journal: Forthcoming, Berlin, Germany, Springer Berlin Heidelberg.

- Doka, K., Filatov, M., Giannakouris, V., Kantere, V., Koziris, N., Mantas, C., Papailiou, N., Papaioannou, V., Tsoumakos, D. (2016). “Optimizing, Planning and Executing Analytics Workflows over Multiple Engines”, International Workshop on Multi-Engine Data AnaLytics (MEDAL).

- Papakonstantinou, N., Zakkak, F., Pratikakis, P. (2016). “Hierarchical Parallel Dynamic Dependence Analysis for Recursively Task-Parallel Programs”, IEEE International Parallel and Distributed Processing Symposium 2016 (IPDPS).

- Scharl, A., Weichselbraun, A., Göbel, M., Rafelsberger, W. and Kamolov, R. (2016). "Scalable Knowledge Extraction and Visualization for Web Intelligence", 49th Hawaii International Conference on System Sciences (HICSS-2016). Kauai, USA: IEEE Press. 3749-3757. BEST PAPER AWARD

- Vandierendonck, H., Arif, M., Nikolopoulos, D.S., de Supinski, B. (2016). “Data Analytics without Map-Reduce”. Submitted to the International Conference on Supercomputing.

- Weichselbraun, A., Scharl, A. and Gindl, S. (2016). "Extracting Opinion Targets from Environmental Web Coverage and Social Media Streams", 49th Hawaii International Conference on System Sciences (HICSS-2016). Kauai, USA: IEEE Press. 1040-1048.

2015

- Arif, M., Vandierendonck, H. (2015). A Case Study of OpenMP Applied to Map/Reduce-Style Applications. 11th International Workshop on OpenMP, IWOMP 2015, Aachen, Germany.

- Doka, K., Tsoumakos, D., Koziris, N. (2015). “Intelligent, Multi-Engine Resource Scheduler for Big Data Analytics Workflows”, European Data Forum 2015 (EDF).

- 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.

- Kantere, V. (2015). “Approximate Querying in Big Heterogeneous Data Sources” IEEE International Congress on Big Data (IEEE BigData): 712-715.

- Kantere, V., Filatov, M. (2015). “A Workflow Model for Adaptive Analytics on Big Data”, IEEE International Congress on Big Data (IEEE BigData): 673-676.

- Kantere, V., Filatov, M. (2015). “A Framework for Big Data Analytics”, In the International C* Conference on Computer Science & Software Engineering (C3S2E): 125-132.

- Kantere, V. (2015). “Datom: Towards Modular Data Management”, IEEE International Conference on Information Reuse and Integration (IEEE IRI): 443-450.

- Kantere, V., Kementsietsidis, A., Orfanoudakis, G., Sellis., T. (2015). “Query Relaxation across Heterogeneous Data Sources”, ACM International Conference on Information and Knowledge Management (CIKM): 473-482.

- Kantere, V., Filatov, M. (2015). “Modelling Processes of Big Data Analytics”, International Conference on Web Information System Engineering (WISE): 309-322.

- 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.

- Papakonstantinou, N., Zakkak, F., Pratikakis, P. (2015). “A Parallel Task Execution Engine” (Submitted).

- Sabou, M., Onder, I., Brasoveanu, A., Scharl, A. (2016. "Towards cross-domain data analytics in tourism: a linked data based approach", Information Technology & Tourism, Berlin (Accepted); Germany, AKA.

- Scharl, A., Herring, D., Rafelsberger, W., Hubmann-Haidvogel, A., Kamolov, R., Fischl, D., Föls, M. and Weichselbraun, A. (2015). “Semantic Systems and Visual Tools to Support Environmental Communication IEEE Systems Journal, Issue: 99, pp 1-10.
- Vandierendonck, H. (2015). Compiler and Runtime Support for Hybrid Static/Dynamic Scheduling. Compilers for Parallel Computing Workshop, CPC-15.

2014

- Katsogridakis, P. and Pratikakis, P. (2014). Hama and Hadoop: “A Performance Comparison”, 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, a poster and a demo in 2015 in top-tier international and European conferences. Moreover, a demo and a research paper submission in a prestigious data management conference is under review. ICCS 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 Modelling, etc. In 2016 submissions will increase, as soon as the first tangible results of WP3 are available.

IMR mostly participates in workshops and seminars where the research and development team is invited to present technical achievements, performance evaluations, use cases and industrial assessments of open source tools for Big Data management. We presented our activities at Big’14, a workshop collocated with WWW-2014. Advanced Big data analytics issues related to our use cases investigated in ASAP have been presented at Digital Tuesday (www.digital-tuesday.net/paris) and at the Digital Luxe Meeting (www.digital-luxe-meeting.fr/#dlm2015).

UniGe has started disseminating the initial research results to international scientific conferences. Overall in we have already published 2 conference papers in 2014 and 6 conference papers in 2015 in venues with excellent reputation and of very wide visibility. We are continuing our work on the research topics of Work Package 5 and we plan to have 4 papers at least submitted in 2016.

webLyzard. A paper on scalability improvements in the WLT knowledge extraction processing pipeline has received a *Best Paper Award* at the *49th Hawaii International Conference on System Sciences* (HICSS-2016). WLT plans to submit a refined and extended version of this paper to a journal outlet, incorporating the Year 2 progress made in WP6. Once the integration of the WLT content knowledge repository and the telecommunications dataset of WIND are complete, we also plan a follow-up publication on the visualization of statistical data for decision support applications – building on an upcoming article to appear in the *Semantic Web Journal*. In terms of target outlets, in addition to the examples listed at the beginning of Section 2.2.1, we will continue to focus on ACM and IEEE journals with a focus on information management and visual analytics.

FORTH disseminates scientific results of the project via research paper conference publications, journal publications, workshop presentations, and poster presentations in related scientific venues. Related areas include High-Performance Computing, Big Data, Distributed Systems, Parallel Programming and Programming Languages. FORTH has organized a thematic session in the HiPEAC Computing Systems Week event in October 2014. FORTH has both published and ongoing work on the ASAP irregular query execution engine and work on programming languages for distributed and parallel execution engines. FORTH expects more publications on these topics during the third year of the project, with one such publication currently under review and plans for at least two more submissions during the first half of 2016.

QUB currently has made a number of informal presentations and will disseminate the scientific results of the project via research paper conference publications, journal publications, workshop presentations, and poster presentations in scientific venues related to high-performance computing and big data. QUB has ongoing work on the ASAP programming language, on high-performance graph analytics and on scalable task scheduling. QUB expects publications on these topics to materialize during the third year of the project and is aiming to publish them at leading conferences in high-performance computing and big data.

1.2.2. Events and Presentations

2016

- Mar 2016: *1st International Workshop on Multi-Engine Data Analytics*[9]

  Dimitrios Tsoumakos (ICCS, technical leader of ASAP) and Verena Kantere (UNIGE) are co-chairing the *1st International Workshop on Multi-Engine Data AnaLytics* (MEDAL 2016), collocated with EDBT 2016 in Bordeaux, France. This workshop, inspired by ASAP, aims to bring together researchers and practitioners from both academia and industry to explore, discuss and possibly redefine the state of the art in big data analytics relative to modelling, methods and tools applied over any part of algorithms and computing infrastructures, as well as use-cases and applications that relate to big data analytics over multi-engine environments. The topics of the workshop include:

  - Modelling of analytics processes: new models and languages to program, represent and execute complex tasks
  - Execution of analytics processes: planning, optimizing and executing complex or multiple workflows especially on dynamic multi-engine and elastic environments
  - Tools for advanced analytics tasks: theoretical and practical development of analytics tasks and operators for regular and irregular computations
  - Visualization of analytics tasks: adaptive, user-friendly and diverse representation of real-time and batch tasks executing over single or multiple runtimes
  - Applications of big data analytics: case-studies, exhibition of application-specific challenges

  The keynote speaker will be Volker Markl, Professor and Chair of the Database Systems and Information Management (DIMA) group at the Technical University Berlin (TU Berlin). His talk will be on topics related to the Stratosphere and Mosaic FP7 ICT projects. The work of the ASAP project will be represented by one research and one demo paper. The workshop will be held on March 15, 2016.

- Jan 2016: Vandierendonck, University of Manchester, School Seminar - The Swan Task Dataflow Scheduler: Design and Applications.

[9] medalworkshop2016.github.io

2015

- Dec 2015: Kantere, Talk at University of Fribourg.
- Nov 2015: Doka, Poster Presentation at the European Data Forum (EDF) 2015, Luxembourg.
- Oct 2015: IMR participation to the Flink forward meeting.[10]
- Oct 2015: Doka, Invited Talk at King Abdullah University of Science & Technology.
- Oct 2015: Scharl, Invited Talk at the Eye on Earth Summit.[11]
- Aug 2015: Presentation of IMR services at the Digital Luxe meeting.[12]
- Jun 2015: Murphy, Soapbox Science Ireland 2015 – Public Outreach Program for Research.
- May 2015: Vandierendonck, Presentation at HiPEAC Computing Systems Week, Oslo, Norway.
- Mar 2015: Presentation of IMR services at the Digital Tuesday.[13]

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 International Summer School on Advanced Computer Architecture and Compilation for High-Performance and Embedded Systems (ACACES 2014).

1.3. Media Coverage

- Feb 2016. Interview with FORTH-researcher on local regional TV station.
- Jun 2015. United Nations Environment Programme (UNEP) commissions MODUL University Vienna | UNEP News Centre
- Nov 2014. Weißes Haus setzt bei neuer Klima-Initiative auf österreichische Web Intelligence | OekoNews AT.
- Jun 2014. Die Meinungs-Erforscher | Computerworld CH.

[10] www.flink-forward.org
[11] www.eoesummit.org
[12] www.digital-luxe-meeting.fr/#dlm2015
[13] www.digital-tuesday.net/paris

1.4. Research Collaboration

ICCS will exploit results from the recently concluded FP7 **CELAR** project,[14] which dealt with the elasticity aspects of applications deployed over cloud infrastructures. Research and tools relevant to profiling, modelling and elasticity decisions customized to big data analytics engines will be used and evolved in the course of ASAP.

IMR closely collaborates with ICSS and UNIGE on a high-level, abstract specification of Big data analytics use cases. We are currently exploring the range of available distributed execution engines that best suit our needs, including Spark and Flink. Regarding the later, our engineers attended the **Flink Forward** meeting in November 2015.[15] We benefit from ASAP activities to finalize our platform architecture by the end of the project.

QUB is a partner of **CACTOS,**[16] which develops tools and methods to acquire and analyse application behaviour and infrastructure performance data. The mined information optimises 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 pursues active research cooperation, identify synergies to maximize impact, and exchange results with other ongoing research projects such as **uComp**,[17] which develops a human computation framework to blend knowledge extraction and crowdsourcing approaches in the tradition of games with a purpose, **DecarboNet**,[18] which builds a collective awareness platform for carbon footprint reductions, **PHEME**,[19] which tackles *veracity* as an emerging challenge of big data research - in addition to Gartner's 3Vs of big data: *volume*, velocity, and variety, and the H2020 Innovation Action **InVid**[20] that detects emerging stories and assess the reliability of user-generated video content (where WLT leads the platform development).

FORTH will exploit results from ASAP in the currently ongoing **Vineyard** project, which deals with the use of accelerators in data centres. We are exploring the use of the ASAP platform as a scheduler and driver for analytics applications that can benefit from GPU and other accelerator hardware; specifically Deep Learning AI. We aim to use this application also in ASAP to demonstrate reusability of the platform in domains outside the application use cases.

[14] www.celarcloud.eu
[15] www.flink-forward.org
[16] www.cactosfp7.eu
[17] www.ucomp.eu
[18] www.decarbonet.eu
[19] www.pheme.eu
[20] www.invid-project.eu

1.5. Software and Data Resources

Software prototypes are made publicly available for download, either from the project Website or the **GitHub.com** public repository at www.github.com/project-asap. The list of currently available components includes:

- **Intelligent Resource Scheduler (IReS)**
  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
- **Workflow**
  Unified, open-source 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
- **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

- **easy 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
- **Media Watch on Climate Change (MWCC)**
  Public showcase of the ASAP dashboard extensions with data export capabilities (T6.1), focusing on climate science and related environmental issues
  www.ecoresearch.net/climate

1.6. Industry-Oriented Materials

As a result of ongoing industry-oriented activities, exploitable results of ASAP are published through the ASAP Website as software packages, presentations and Web documents. At the same time, the industrial partners of the ASAP consortium seek synergies and emphasize commonalities in industry-oriented presentations. We pursue a differentiated approach vis-à-vis specific sectors, depending on which organization is most skilled to address this sector. The presentations aim to convince early adopters of the benefits and opportunities that ASAP can offer.

2. 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.

2.1. 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. One of the main use cases being explored during the project is a classification of structured data extracted from web sources. The use case presents some generic aspects which are worth exploring to evaluate the ability of the ASAP platform to adapt to various workloads and environments: data flows through a sequence of op-

erators, is used to build classification models, and must be continuously processed and adapted. We are using an actual Mignify service for this evaluation, with data continuously extracted from Marketplaces, classified and organized in large catalogues representative of e-Business activities all over the world.

We plan to investigate several project, including a mere centralized implementation for small sources that do not require a full-fledged distributed execution engine, and two concurrent platform operating at web scale, namely Spark and Flink. Collaboration with expert partners in distributed systems and data processing issues help 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**. We plan to publish the data extraction and classification service by the end of the project, thereby improving the attractiveness and visibility of Mignify.

Another direction of R&D that directly contributes to our platform is related to system aspects. Currently, all our services are hosted in the Mignify **data centre.** 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 catalogue of operators, would highly leverage the current potential of Mignify. ASAP helps us design and test such a cost-based approach.

2.2. WIND Telecomunicazioni

WIND contributes to communicating the project results 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 extensive 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 large scale data analysis aiming at trend discovery and smart decision making.

WIND plans on evaluating the effect of the **tourist trend 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). These are some details on how ASAP with its specific orientation can offer innovative implications to industrial performance.

The objective of the project is to develop an innovative forecasting model of **tourist flows** linked to events and places, combining the analysis of traffic data from the telephone network with social media content. 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 science** - 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.

In order to better define such services WIND is in contact with the Mobility Agency of the City of Rome. The collaboration with this entity will allow in the future a practical application of the ASAP framework and of the results coming from the tests done with the 1st ASAP Telecommunication Data Analytics (TDA) application prototype in the beginning of 2016. The services targeted as a result of this collaboration will be mainly on the discovery of people's preferred activities in specific touristic areas in the City of Rome (POI - Points of Interest).

Finally, from the point of view of the company, the ASAP activity allows to test the 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.

2.3. webLyzard technology

In terms of visibility 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 centres** in Europe and the United States, as well as 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 - 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) or international organizations such as the *World Bank* and the *United Nations Environment Programme* (UNEP) will also be beneficial for disseminating ASAP results to a wide international audience. The 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.

In *Year 1*, this included the **U.S. Climate Resilience Toolkit** developed in response to President Obama's Climate Action Plan. webLyzard provided the Toolkit's search function to help visitors quickly locate the most relevant content across U.S. federal government's Websites. The visualization modules developed in T6.1 of WP6 have been integrated into the analytics view of this application, which enables communication experts at the National Oceanic & Atmospheric Administration (NOAA) to monitor Web content streams and continuously improve the CRT **knowledge repository**.

In *Year 2 of ASAP*, the work has concentrated on (i) modularizing the platform and providing a REST API (Application Programming Interface), (ii) integrating and visualizing structured and unstructured use case data to enable joint exploitation activities together with the other ASAP industry partners, and (iv) leveraging the dissemination potential from international collaborations.

- The functionality of the **REST API** includes the ability to upload crawled content, to annotate this content and provide search results via data services, and to embed individual visualization modules into third-party Web applications.
- The achieved integration and visualization of structured and structured content, as demonstrated through the joint processing of **telecommunications data** (WIND) and **Web content metrics** (webLyzard, IMR), paves the way for a wide range of exploitation activities - e.g. **business intelligence** services that relate actual human behaviour (e.g., phone calls and text messages during a public event) with peaks of online coverage and the aggregated perceptions of stakeholder groups
- At the same time, the developed technologies continued to support external collaborations, for example joint projects with the *NOAA Climate Program Office* (including a new ambient search initiative to commence in Q2-2016), and hosting the new Web intelligence component of the United Nations' **UNEP Live Platform**.
- Early showcases such as the *U.S. Climate Resilience Toolkit,* nominated for a **Webby Award 2015**, demonstrate technological leadership and 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**, 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 our products and services, but also to point towards the ASAP project as a driver of innovation behind some of the underlying core technologies.[21]

[21] www.eepurl.com/bbOmkT

In *Year 3* of ASAP, we plan to continue and intensify these activities to promote ASAP in conjunction with the webLyzard Web intelligence platform. Its portfolio of *visualization services* are a core value proposition of webLyzard. The availability of a comprehensive **REST API** that serves as an interface to the repository as well as access to these visualization services will unlock new exploitation potential, following a *Visualization-as-a-Service* (VaaS) approach, or possibly even a *Container-as-a-Service* (CaaS) approach when deployed in conjunction with the Docker platform.

Both use cases present opportunities to introduce products soon after the successful completion of the ASAP project, which will include specific Web intelligence offers for (i) **telecommunications** and **tourism** companies, initially targeting the Italian and Austrian markets (WP9), and (ii) large **B2C brands** operating *internationally* and using *e-commerce marketplaces* as distribution channels for their products (WP8).

Appendix A: ASAP Fact Sheet

ASAP Research Project - A Scalable Analytics Platform

www.asap-fp.eu • @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).

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) 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

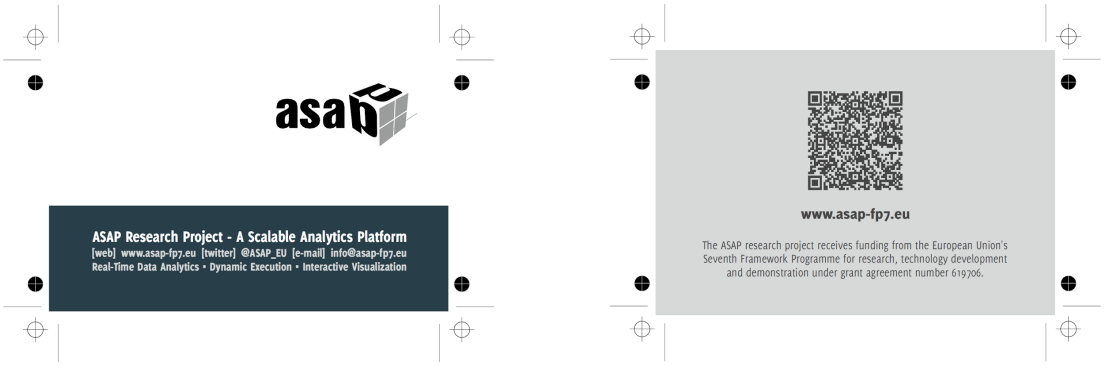
- Foundation for Research and Technology – Hellas
  www.forth.gr
- Université de Genève
  www.unige.ch
- Institute of Communication and Computer Systems
  www.iccs.gr/eng
- Queen's University Belfast
  www.qub.ac.uk
- Internet Memory Research
  www.mignify.com
- WIND Telecomunicazioni
  www.wind.it
-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

Appendix B: Promotional Materials

Business Cards

WEBLYZARD organized the printing of 4,000 **business cards** using the corporate identity established through the WordPress theme of the ASAP Website. 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.



T-Shirts

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.

FP7 Project ASAP
Adaptable Scalable Analytics Platform

End of ASAP D10.5
Yearly Report for Exploitation and Dissemination of Foreground

WP 10 - Exploitation and Dissemination

Nature: Report
Dissemination: Public