FP7 Project ASAP
Adaptable Scalable Analytics Platform

ASAP D10.6
Final Report for Exploitation and Dissemination of Foreground

WP 10 - Exploitation and Dissemination

Nature: Report
Dissemination: Public

Version History

Version	Date	Author	Comments
0.1	17 Dec 2016	A. Scharl	Initial Version based on D10.5
0.2	18 Jan 2017	A. Scharl	Major Revision
0.3x		All Partners	Various Content Updates
0.5	28 Feb 2017	A. Brasoveanu	Updates and Revision
1.0	08 Mar 2017	A. Scharl	Minor Revision

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 during the three years of the ASAP project, describes future exploitation activities planned after the project has ended, and outlines how the project results have been made available to the **European industry**. This document has been continuously updated throughout the project (D10.4-D10.6).

All partners have played 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** have formed a key part of the dissemination strategy towards non-scientific users to provide **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. The consortium is committed to keep the Website online and up-to-date for at least three years after the end of the project. The Website is being supported by the ASAP **Twitter channel**,[2] which also reports new results as they become available. The scientific results of the project have been or are currently being submitted to major international **conferences** and **journals**, and presented in relevant national and European **concertation events**.

To promote maximum use and dissemination, major technology components from ASAP have been 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 were 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 understands 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 supports 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.

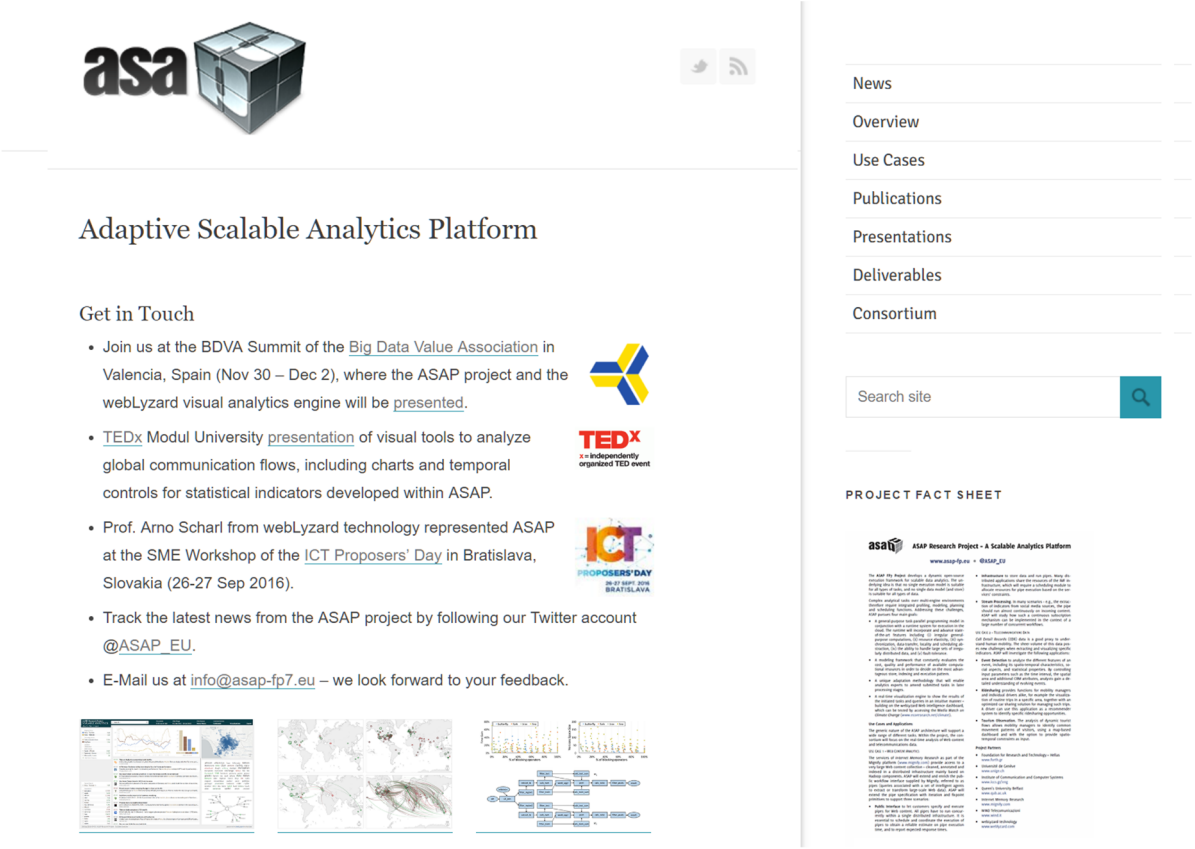
1.1.2 Project Website

The project maintains a public **Website** at **www.asap-fp7.eu**, allowing distributed editing among consortium members. Following an initial static HTML site, a dynamic version 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 independent editing of document by consortium partners.

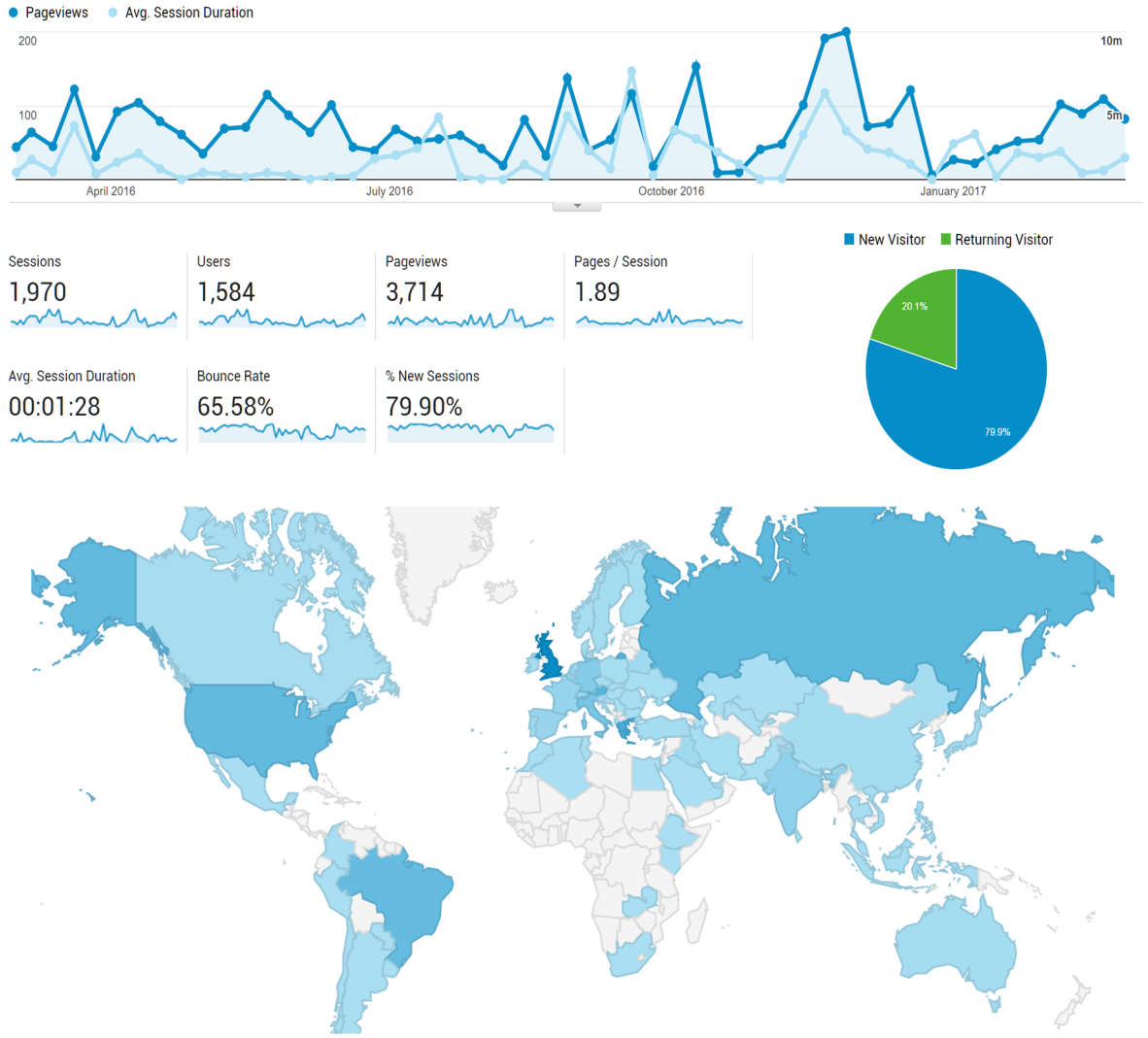
The consortium is committed to maintain the site for at least three years after the end of the project to ensure that project results remain visible and accessible. Throughout the project, the site has been updated regularly, as progress was made and new output became available. The content of the Website includes:

- Project overview and objectives;
- Project news and presentations by consortium members;
- Use case descriptions (Web content and telecommunications data analytics);
- Downloadable project outputs including deliverables, software and data resources, and a list of scientific publications;
- Profiles and roles of project partners;



An analysis of the log files shows that the Website has attracted **1,584 users** requesting **3,714 pageviews** between March 2016 and February 2017. Efforts to improve the structure and layout of the ASAP Website are reflected in an average *session duration* of 1:28 (up 265 % from Year 2), and a rather low bounce rate of 65% (20% lower as compared to Year 2). In terms of the users' **geographic origin**, the leading countries were the United Kingdom (17.7%), Greece (10.4%), Austria (9.1%), Russia (8.3%), and the United States (7.8%).

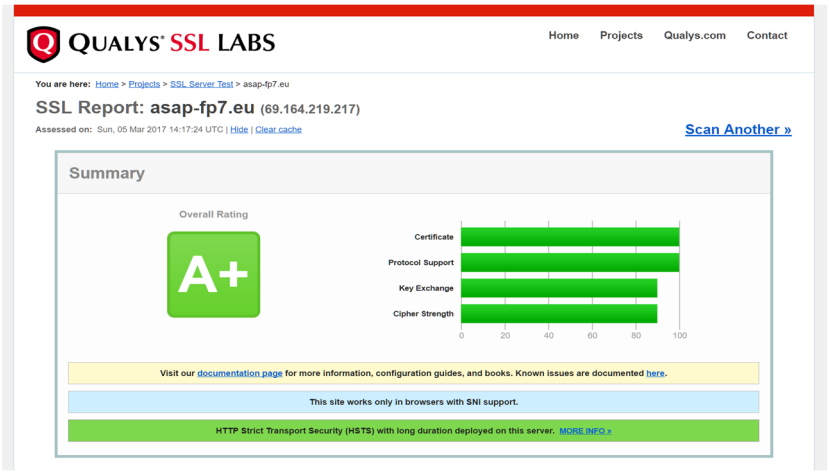
This Website functions as a project dissemination and documentation tool and single access point that references other ASAP **applications** and **services**. In addition to documenting the scientific output of the project, examples and descriptions outline how the information services can be used by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

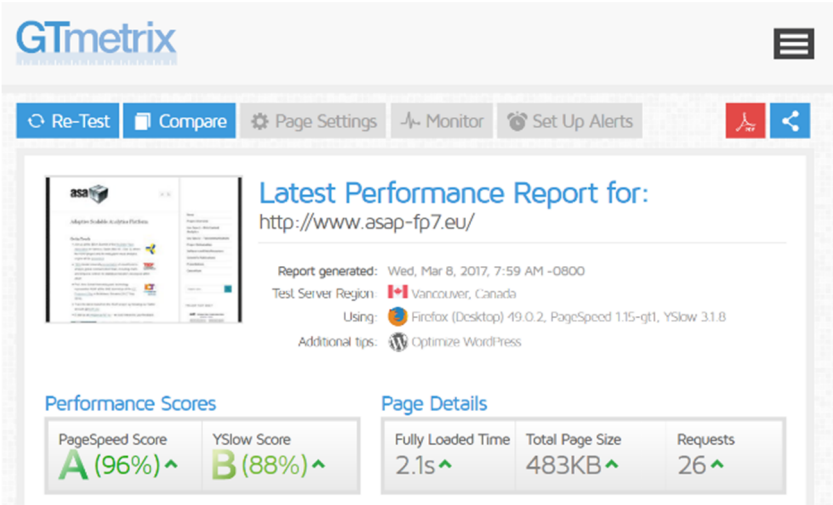


During the third year of the project, the Web site has been improved not only in terms of content and presentation, but also in terms of technical optimization. We have changed the **certificate authority**,[5] tested and deployed a number of **Wordpress plugins** (search engine optimization, concatenation of CSS and JavaScript code, improved lossless image compression, etc.) and adopted the **Content Delivery Network** approach of the ASAP Dashboard (D6.4) for asap-fp7.eu as well. This not only increases the overall user experience, but also the Website's Google ranking and thereby supports the long-term dissemination and exploitation goals of ASAP.

[5] www.letsencrypt.org

The impact of these optimization efforts is documented by an **A+ Rating** of the *Qualsys SSL Server Test,*[6] as well as a **96% PageSpeed** score and an **88% YSlow** Score as measured by the *GTmetrix* online service.[7]



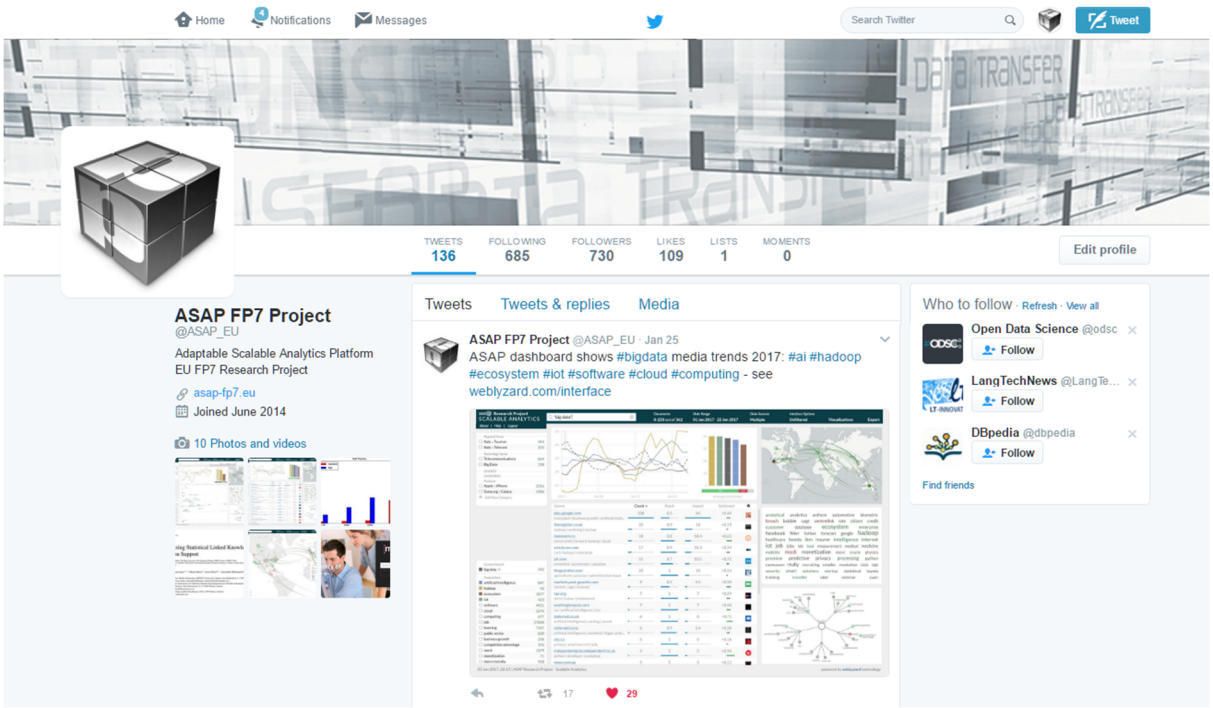


1.1.3 Social Media Channels

Posting project-related information to social media platforms increases the visibility of the ASAP project and promotes the uptake of its results. ASAP has established a Twitter account (**@ASAP_EU**), giving the project partners a channel to disseminate publications, announce events, or promote showcases and public presentations of the project (either the account name or the hash tag **#asap-fp7** is used to tag posts and published material).

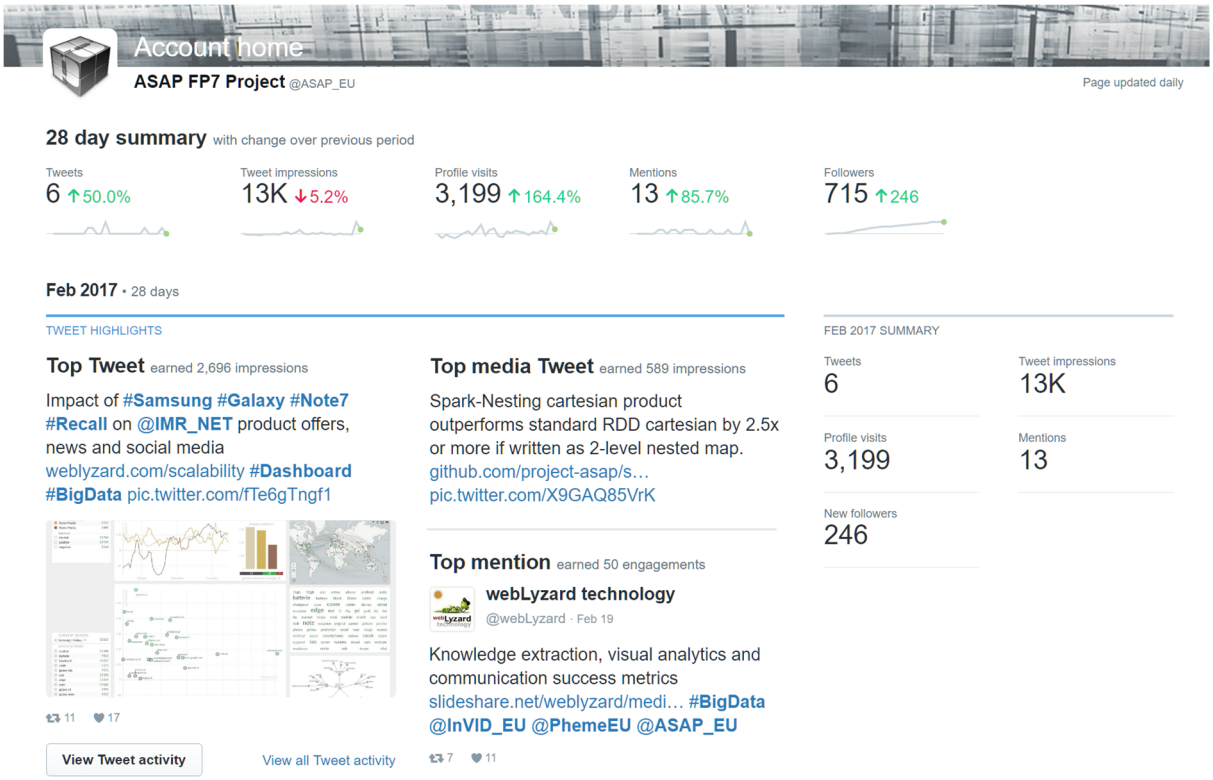
As of February 2017, the Twitter account has attracted a total of **715 followers** (up 420% compared to February 2016). Project members also disseminate ASAP results in additional social media channels such as Facebook, Google+ and LinkedIn.

[6] www.ssllabs.com/ssltest
[7] www.gtmetrix.com



1.1.4 Twitter Impact Metrics

The following diagrams reflect the **impact** of the project's social media activities in terms of organic growth of the @ASAP_EU audience, including the summary of the **final month** (February 2017), overall follower **demographics**, as well as the number of impressions and engagements of the **top tweets** in the last six weeks of the project.



Audience insights

ASAP FP7 Project @ASAP_EU

Interests

Your followers · All Twitter Users

Interest name	Comparison	Difference
Technology	88% / 24%	64% more
Tech news	85% / 22%	63% more
Business news and general info	71% / 31%	40% more
Business and news	66% / 47%	19% more

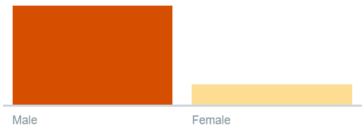
Country

Country name	% of audience
United States	23%
United Kingdom	10%
Germany	7%
Spain	6%
India	6%
Italy	5%
Austria	5%

Your current follower audience size is 729

That's 249 more than the same time 30 days ago. You've gained around 8 new followers per day

Gender



Languages

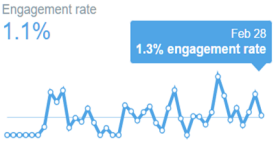
Tweets | Top Tweets | Tweets and replies | Promoted

	Impressions	Engagements	Engagement rate	
ASAP FP7 Project @ASAP_EU · Jan 25 ASAP dashboard shows **#bigdata** media trends 2017: **#ai #hadoop #ecosystem #iot #software #cloud #computing** - see weblyzard.com/interface pic.twitter.com/cS7PRl0ekR View Tweet activity	4,350	163	3.7%	Promote
ASAP FP7 Project @ASAP_EU · Feb 27 Impact of **#Samsung #Galaxy #Note7 #Recall** on **@IMR_NET** product offers, news and social media weblyzard.com/scalability **#Dashboard #BigData** pic.twitter.com/fTe6gTngf1 View Tweet activity	4,170	118	2.8%	Promote
ASAP FP7 Project @ASAP_EU · Feb 11 **.@ASAP_EU** will be presented by **@IMR_NET** at the Web Summit 2017 in Lisbon, Portugal websummit.net View Tweet activity	866	10	1.2%	Promote
ASAP FP7 Project @ASAP_EU · Feb 10 Beta version of the new Bomerce service, supported by the ASAP classification workflow bomerce.io **#shopping #b2c** View Tweet activity	638	8	1.3%	Promote
ASAP FP7 Project @ASAP_EU · Feb 14 Oh, and we're 10% faster on small datasets too. :-) github.com/project-asap/s… View Tweet activity	637	18	2.8%	Promote
ASAP FP7 Project @ASAP_EU · Feb 14 Telecom analytics app used to handle 1 day of data per batch, max. Now scaled up to 2 months per batch, under 4 hrs. github.com/project-asap/s… View Tweet activity	603	12	2.0%	Promote

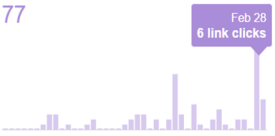
Engagements

Showing 42 days with daily frequency

Engagement rate
1.1%

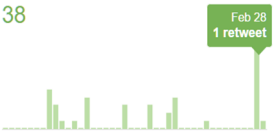


Link clicks
77



On average, you earned **2 link clicks** per day

Retweets
38



On average, you earned **1 Retweets** per day

Likes
71

On average, you earned **2 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 project's results 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.[8]

In order to ensure the widest possible audience, ASAP used **press releases, newsletters** and other types of **dissemination activities** aimed at the academic and industrial sectors (e.g., delivery of tutorials, industry-oriented workshops, etc.). In the third year of the project, members of the ASAP consortium organized the *1st International Workshop on Multi-Engine Data AnaLytics* (MEDAL 2016)[9] in Bordeaux, for example, and presented ASAP results at the *ICT Proposers' Day*[10] in Bratislava and the *BDVA Summit in Valencia*.[11]

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 **37 refereed publications and 6 more currently under review**:

- 4 journal papers,
- 1 book chapter
- 24 conference papers,
- 6 workshop papers,
- 1 summer school paper, and
- 1 conference demonstration.

The **annual number** of publications has remained stable at 13 with 6 paper currently being under reviewer and further papers under preparation.

[8] www.github.com/project-asap

[9] medalworkshop2016.github.io - the MEDAL workshop was co-located with the *19th International Conference on Extending Database Technology* (EDBT-2016)

[10] ec.europa.eu/digital-single-market/en/ict-proposers-day-2016

[11] www.bdva.eu/?q=valencia-summit

Published or Accepted Refereed Publications

2017

- Brasoveanu, A.M.P., Sabou, M., Scharl, A., Hubmann-Haidvogel, A. and Fischl, D. (2017). “Visualizing Statistical Linked Knowledge Sources for Decision Support”, Semantic Web Journal, (8)1, pp. 113-137.
- Filatov, M. and Kantere, V. (2017). “Multi-workflow optimization in PAW”, in EDBT, March 19-23, 2017, (accepted).
- Filatov, M. and Kantere, V. (2017). “Tutorial on Data Analytics in Multi-Engine Environments”, in DASFAA, March 27-30, 2017, Suzhou, China (accepted).
- Weichselbraun, A., Gindl, S., Fischer, F., Vakulenko, S. and Scharl, A. (2017). “Aspect-Based Extraction and Analysis of Affective Knowledge from Social Media Stream”, IEEE Intelligent Systems: Forthcoming.

2016

- Arif, M., Vandierendonck, H., Nikolopoulos, D.S. and de Supinski, B.R. (2016). "A Scalable and Composable Map-Reduce System", in Advances in Software and Hardware for Big Data to Knowledge Discovery (ASH) in conjunction with IEEE International Conference on Big Data.
- Doka, K., Filatov, M., Giannakouris, V., Kantere, V., Koziris, N., Mantas, C., Papailiou, N., Papaioannou, V. and Tsoumakos, D. (2016). “Optimizing, Planning and Executing Analytics Workflows over Multiple Engines”, International Workshop on Multi-Engine Data AnaLytics (MEDAL).
- Doka, K., Papailiou, N., Giannakouris, V., Tsoumakos, D. and Koziris, N. (2016). “Mix 'n' Match Multi-Engine Analytics”, in Proceedings of the 2016 IEEE International Conference on Big Data (BigData 2016), December 5-8, 2016, Washington DC, USA.
- Filatov, M. and Kantere, V. (2016). “PAW: A Platform for Analytics Workflows”, in International Conference on Extending Database Technology, March 15-18, 2016, Bordeaux, France.
- Giannakouris, V, Papailiou, N., Tsoumakos, D. and Koziris, N. (2016). “MuSQLE: Distributed SQL Query Execution Over Multiple Engine Environments”, in Proceedings of the 2016 IEEE International Conference on Big Data (BigData 2016), December 5-8, 2016, Washington DC, USA.
- Mytilinis, I., Tsoumakos, D. and N. Koziris, N. (2016). “Distributed Wavelet Thresholding for Maximum Error Metrics”, in Proceedings of the 2016 ACM SIGMOD/PODS International Conference on Management of Data, June 26 - July 1, 2016, San Francisco, USA.

- Papailiou, N., Doka, K., Giannakouris, V., Papaioannou, V., Tsoumakos, D. and Koziris, N. (2016). “Robust and Adaptive Multi-Engine Analytics using IReS (demo paper)”, in Proceedings of the 10th International Workshop on Enabling Real-Time Business Intelligence (BIRTE 2016), in conjunction with the VLDB 2016 Conference, 2016, New Delhi, India.

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

- Sabou, M., Onder, I., Brasoveanu, A. and Scharl, A. (2016). "Towards cross-domain data analytics in tourism: a linked data based approach", Information Technology & Tourism, Vol. 16, Issue 1, pp 71–101.

- Scharl, A., Lalicic, L. and Önder, I. (2016). "Tourism Intelligence and Visual Media Analytics for Destination Management Organizations", Analytics in Smart Tourism Design - Concepts and Methods. Eds. Z. Xiang and D.R. Fesenmaier. Cham: Springer. 165-178.

- 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, pp. 3749-3757. BEST PAPER AWARD

- Sun, J. (2016). “Student Research Poster: A Scalable General-Purpose System for Large-Scale Graph Analytics.”, in PACT '16: Proceedings of the 25th international conference on Parallel architectures and compilation, pp. 456-456.

- Vandierendonck, H., Murphy, K., Arif, M., Sun, J. and Nikolopoulos, D.S. (2016). “Operator and Workflow Optimization for High-Performance Analytics”, International Workshop on Multi-Engine Data AnaLytics (MEDAL).

- Vandierendonck, H., Murphy, K., Arif, M. and Nikolopoulos, D.S (2016). “HPTA: High-Performance Text Analytics”, in Proceedings of the 2016 IEEE International Conference on Big Data (BigData 2016), December 5-8, 2016, Washington DC, USA.

- 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, pp. 1040-1048.

2015

- Arif, M. and 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. and 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”, in Proceedings of the 2015 ACM SIGMOD/PODS International Conference on Management of Data (SIGMOD'15, Demo track), pp. 1451-1456.

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

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

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

- Kantere, V. and Filatov, M. (2015). “Towards Adaptive Analytics on Big Data Sources”, International Conference on Data Mining, Internet Computing, and Big Data (BigData2015), pp. 84-94

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

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

- Kantere, V. and Filatov, M. (2015). “Modelling Processes of Big Data Analytics”, International Conference on Web Information System Engineering (WISE), pp. 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", in IEEE International Conference on Cloud Engineering (IEEE IC2E 2015), pp. 201-206.

- Papailiou, N., Tsoumakos, D., Karras, P. and Koziris, N. (2015). "Graph-Aware, Workload-Adaptive SPARQL Query Caching", in Proceedings of Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, pp.1777-1792.

- 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: 99, pp. 1-10.

- Vandierendonck, H. (2015). “Efficiently Scheduling Task Dataflow Parallelism: A Comparison Between Swan and QUARK”, in Proceedings of the Exascale Applications and Software Conference. pp. 36-41.

- Vandierendonck, H. (2015). “Compiler and Runtime Support for Hybrid Static/Dynamic Scheduling”, Compilers for Parallel Computing Workshop, CPC-15.

2014

- 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).

- 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).

Submitted Papers Currently under Review

- 6 Conference papers have been submitted and are currently under review

Trade Publications

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

Future Publication Plans

FORTH. During the third year of the project FORTH has published a paper at IPDPS, submitted a paper for review at EuroPar, and plans to submit a paper at HotCloud 2017 in March. FORTH has also disseminated scientific results of the project in one invited presentation at Credit Suisse, a news release to Greek media in February, a page-entry news release to the HiPEAC newsletter for March, and plans an extended news release to Greek media upon completion of the project review. Finally, FORTH has submitted input to open issues posted on the Spark repository regarding the support for nested and recursive computations, to increase traffic in the Spark-Nesting repository. FORTH plans to continue making heavy use of the Spark-Nesting module in analytics research in the future. We use Spark-Nesting to implement a library of graph computations not currently supported by GraphX, albeit often requested by the community, targeting social-network analytics computations (network centrality, influencer detection, shortest-path enumeration, etc). We expect publications from these results to be submitted to social-network analysis venues within 2017; although this analysis may have limited overlap with the immediate objectives of ASAP, these implementations would not be possible without support for recursive analytics computations.

UNIGE has started disseminating the initial research results to international scientific conferences. Overall, 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. In 2016 we published 3 more conference papers and 1 paper in the 1st International Workshop on Multi-Engine Data AnaLytics (MEDAL), held in conjunction with EDBT 2016, which we organized together with other ASAP partners. We also presented a tutorial on Data Analytics in Multi-Engine Environments at the International Conference on Data Analytics and Management in Data Intensive Domain (DAMDID). UNIGE is continuing disseminating the results obtained during the ASAP project, and we have already accepted two conference papers for 2017, in EDBT and ICDCS, as well as a second presentation of the tutorial on Data Analytics in DASFAA.

ICCS published 4 research papers during the last year of the project: 3 of them in top-tier international conferences in the field of Big Data (IEEE BigData) and Data Management (ACM SIGMOD) and 1 in an international workshop (BIRTE), collocated with one of the most prestigious conferences in the field of Databases (VLDB). Moreover, ICCS presented IReS at the graduate course “Distributed Systems” of the Department of Informatics & Telecommunications of the University of Athens. ICCS intends to further disseminate project results even after the end of the project 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.

QUB. In Year 3 of the project, QUB has made a number of informal presentations, including a seminar at the Summer School on Advances in Programming Languages and a presentation at the EU HiPEAC Network (a former Network-of-Excellence). QUB has disseminated the scientific results of the project via research paper publications at workshops (2) and conferences (1) and poster presentations (1) in scientific venues related to high-performance computing and big data. QUB has published core work on the ASAP programming language and its application to map/reduce programs, to text analytics and to high-performance graph analytics. Further publications on graph analytics (one paper submitted to International Conference on Supercomputing (ICS) and one in preparation for International Conference on Parallel Processing (ICPP)), on the scheduler (submitted to International Conference on Supercomputing (ICS)) and on the optimizations to TF/IDF in Spark (venue TBC) are in progress.

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. During the last year of the project, we participated to several fairs and professional conferences to present our new services centered of eMarketPlace analysis, for which a part of our processes have been evaluated with ASAP. In particular, the Presence service was presented at Web Summit 2016 in Lisbon[12], and the Bomerce service will be publicly released in March 2017.

webLyzard. The visualization methods developed within ASAP have contributed to a number of scientific articles published in Year 3, including a chapter on *tourism intelligence* in an edited Springer volume, a journal article on *linked data for cross-domain decision making*, and an upcoming article on *aspect-based knowledge extraction* in IEEE Intelligent Systems. WLT plans to leverage the developed tools as part of a dedicated tourism intelligence platform, and submit an analysis of the resulting dataset to a journal outlet - incorporating WP6 progress on geospatial projections made in Year 3. Using the integrated dataset of WLT social media content and WIND telecommunications data, we also plan a follow-up publication on the visualization of statistical data for decision support applications. In terms of target outlets, in addition to the examples listed at the beginning of Section 2.2.1, we will focus on leading information systems and tourism journals.

[12] websummit.com

1.2.2 Events and Presentations

2017

- Mar 2017: Filatov, Tutorial on Data Analytics in Multi-Engine Environments at 22nd International Conference on Database Systems for Advanced Applications[13] in Suzhou, China
- Feb 2017: Spada, Round Table on Big Data Research at BigData Committee of ISTAT (Italian Statistical Bureau) in Rome, Italy

2016

- Dec 2016: Arif, Workshop "A Scalable and Composable Map-Reduce System" at 3rd Workshop on Advances in Software and Hardware for Big Data to Knowledge Discovery (ASH)[14] in Washington D.C., USA
- Dec 2016: Doka, Presentation "Mix 'n' Match Multi-Engine Analytics" at IEEE International Conference in BigData 2016[15] in Washington D.C., USA
- Dec 2016: Giannakouris, Presentation "MuSQLE: Distributed SQL Query Execution Over Multiple Engine Environments" at at IEEE International Conference in BigData 2016[16] in Washington D.C., USA
- Dec 2016: Murphy, Presentation "HPTA: High-Performance Text Analytics" at IEEE International Conference in BigData 2016[17] in Washington D.C., USA
- Dec 2016: Trasarti, Presentation "Analyzing Wind Data" at ISTAT (Italian Statistical Bureau) in Pisa, Italy
- Dec 2016: Scharl, Presentation "Cross-Lingual Knowledge Extraction and Visual Analyticsl"[18] at BDVA Big Data Summit 2016, Valencia, Spain
- Nov 2016: Vandierendonck, Invited Talk "In-Memory Data Management for High-Performance Analytics" at HiPEAC Computing Systems Week, Thematic Session on Systems Support for Big Data Applications[19] in Dublin, Ireland
- Oct 2016: Filatov, Invited Talk "Data Analytics in Multi-Engine Environments" at Data Analytics and Management in Data Intensive Domains;[20] Ershovo, Russia

[13] ada.suda.edu.cn/dasfaa2017/Conference#Tutorials
[14] www.cecsresearch.org/ASH/
[15] cci.drexel.edu/bigdata/bigdata2016/
[16] cci.drexel.edu/bigdata/bigdata2016/
[17] cci.drexel.edu/bigdata/bigdata2016/
[18] www.slideshare.net/weblyzard/bdva-summit-valencia
[19] www.hipeac.net/csw/2016/dublin/
[20] damdid2016.frccsc.ru/en/tutorials.html

- Oct 2016: Scharl, Presentation “Analyzing the Digital Talk: Visual Tools for Exploring Global Communication Flows”[21] at TEDx MODUL University Vienna in Vienna, Austria
- Oct 2016: Trasarti, Presentation “Big Data application: a real case of study” at the University of Pisa, Big Data Analytics Business Informatics Course[22]
- Oct 2016: Trasarti, Presentation of Mobility Applications, Pisa Internet Festival[23]
- Sep 2016: Scharl, Presentation of ASAP results at ICT Proposers Day in Bratislava, Slovakia
- Jul 2016: Doka, Presentation of the IReS platform at the University of Athens, Graduate Course Distributed Systems
- Jun 2016: Scharl, UNEP Live Web Intelligence - Dashboard Presentation[24] at Eye on Earth Community, online
- May 2016: Papailiou, Demo of IReS at the BIRTE'16 Workshop (co-located with VLDB'16) in New Delhi, India
- May 2016: Scharl, Live Demo at the Plenary Session of the Science Policy Forum at United Nations Environment Assembly (UNEA-2), Science Policy Forum[25] in Nairobi, Kenya
- May 2016: Scharl, Presentation “UNEP Live Web Intelligence - A Big Data Approach to Analyzing and Visualizing Stakeholder Communication from Online Sources” at United Nations Environment Assembly (UNEA-2), Sustainable Innovation Expo[26] in Nairobi, Kenya
- Apr 2016: Vandierendonck, Invited Talk “Hybrid Static/dynamic Scheduling and the Quest for Extremely Fine-Grain Parallelism” at SIAM Conference on Parallel Processing in Paris, France
- Mar 2016: *1st International Workshop on Multi-Engine Data Analytics.*[27] Dimitrios Tsoumakos (ICCS, technical ASAP lead) and Verena Kantere (UNIGE) co-chaired the *1st International Workshop on Multi-Engine Data AnaLytics* (MEDAL 2016), collocated with EDBT 2016 in Bordeaux, France.
- Jan 2016: Vandierendonck, University of Manchester, School Seminar - The Swan Task Dataflow Scheduler: Design and Applications.

[21] www.youtube.com/watch?v=_P3lAFR3LRQ
[22] didawiki.di.unipi.it/doku.php/bigdataanalytics/bda/start
[23] www.internetfestival.it/
[24] www.eoesummit.org/community/about-the-eoe-community/
[25] www.unep.org/unea/calendar/unea-science-policy-forum
[26] www.sustainableinnovationexpo.org/agenda
[27] 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.[28]
- Oct 2015: Doka, Invited Talk at King Abdullah University of Science & Technology.
- Oct 2015: Scharl, Invited Talk at the Eye on Earth Summit.[29]
- Aug 2015: Presentation of IMR services at the Digital Luxe meeting.[30]
- 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.[31]
- **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).

[28] www.flink-forward.org
[29] www.eoesummit.org
[30] www.digital-luxe-meeting.fr/#dlm2015
[31] digital-tuesday.net/paris

1.3 Media Coverage, Communication and Press Releases

- 2017 (planned): News release to Greek Media.
- Nov 2016: Project Result Summary submitted to HiPEAC newsletter | In-print newsletter circulated in HiPEAC conferences.
- Nov 2016: Der US-Wahlkampf im virtuellen Showdown | Press Release. prd.at/newsroom-kunden/der-us-wahlkampf-im-virtuellen-showdown
- Jun 2016: The Virtual Road to the White House | Press Release. prd.at/en/newsroom-clients/the-virtual-road-to-the-white-house-us-election-2016-web-monitor-tracks-online-perceptions-of-the-presidential-race
- Jun 2016: Big Data for Sustainability | UNEP Press Release. uneplive.unep.org/media/docs/web_intelligence/19_May_MODULUniv_PR.pd
- May 2016: UN nutzt Wiener Know-How für Big-Data und Umwelt Media Coverage by | Media Coverage by Trendkraft.de. trendkraft.de/wissenschaft-forschung-technik/un-nutzt-wiener-know-how-fuer-big-data-umwelt/
- Feb 2016: Interview with FORTH-researcher on local regional TV station.
- Sep 2015: Intelligence is in the Details | webLyzard Newsletter #27. www.eepurl.com/bxfe3j
- Jun 2015: United Nations Environment Programme (UNEP) commissions MODUL University Vienna | UNEP News Centre. web.unep.org/newscentre/united-nations-environment-programme-unep-commissions-modul-university-vienna
- Jan 2015: Business Intelligence | www.weblyzard.com/business-intelligence
- Nov 2014: Weißes Haus setzt bei neuer Klima-Initiative auf österreichische Web Intelligence | Media Coverage by OekoNews AT.
- Nov 2014: U.S. Climate Resilience Toolkit | webLyzard Newsletter #18 www.eepurl.com/7BT4b
- Jun 2014: Die Meinungs-Erforscher | Swiss CIO Magazine Article.
- May 2014: Radical Scalability | webLyzard Newsletter #12 www.eepurl.com/Tsvln
- May 2014: Scalable Analytics | www.weblyzard.com/scalability

1.4 Research Collaboration

FORTH has started using results from ASAP in the currently ongoing Vineyard project, which deals with the use of accelerators in data centres. We are encoding and re-coding GPU-based machine learning algorithms using the Spark-Nesting scheduler. We are also planning to use the whole ASAP platform to drive benchmarks and applications in the Vineyard project once the GPU virtualization prototyping reaches beta. Moreover, we currently participate in a new consortium formed to target embedded applications with a data center component, where we plan to port ASAP technology to optimize and schedule alternative implementations of operators, and dynamically choose between local/embedded vs cloud-backed operators in car automation. Finally, we are investigating the possibility of applying ASAP results to the streaming computations, and are currently in contact with financial, trading, and insurance institutions with such applications, in the hope that this will lead to a possible continuation of the analytics research into the stream processing domain.

UNIGE intends to exploit the results from ASAP in a new project that is titled "Deployment Optimization for Big Data Applications on Hybrid Large-Scale Computing Infrastructures" that will be funded by the Swiss National Science Foundation. This project is in collaboration with the Swiss Institute of Bioinformatics (SIB) and the European Organization for Nuclear Research (CERN), which own huge multi-engine environments. UNIGE will have the leading role in this new project, and will produce a deeper understanding of how to optimize the deployment of Big Data Applications on hybrid large-scale infrastructures. Using a novel combination of Big Data analytics and modeling results, our aim is to improve the performance of of the infrastructures of the Worldwide LHC Computing Grid (WLCG) at CERN in high-energy physics, Vital-IT (part of the Swiss Institute for Bioinformatics (SIB)).

ICCS will exploit results from ASAP in the H2020 ACTiCLOUD[32] and SELIS[33] projects, which recently kicked-off. SELIS aims to create a Shared European Logistics Information Space where different logistics stakeholder communities will share and use information to manage their business in the most efficient, safe and green way whilst maintaining absolute control of their own data. SELIS will be powered by Big Data Analytics, feeding green logistics Decision Support Applications. ICCS has undertaken the role of designing and implementing the Big Data analytics platform and will thus use and evolve the relevant ASAP tools, customizing them to the needs of the logistics domain. ACTiCLOUD aims to unify the physical resources of cloud sites in order to offer better utilization and scalability to resource-demanding applications. Thus, it will be able to rely on the expertise gained through ASAP in the fields of resource provisioning and application modelling.

[32] acticloud.eu/
[33] www.selisproject.eu/

QUB is a partner of **VINEYARD,**[34] which develops a programming framework and runtime system for scheduling applications in a data centre using accelerators such as Graphical Processing Units (GPU) and Field Programmable Gate Arrays (FPGA). The understanding of code structure and optimisation opportunities gained in ASAP will be directly relevant to VINEYARD. QUB will moreover use the benchmark implementations developed in ASAP using the Swan language in VINEYARD. QUB will furthermore utilise its knowledge developed in Spark in the **HPDCJ**[35] project, where it is investigating garbage collection strategies for data-intensive applications.
IMR will continue to explore distributed machine-learning algorithm for continuous processing of eMarket data. We plan to introduce in our workflows distributed classification and distributed recommendation, executed thanks to the Flink distributed engine. Work with partners of the H2020 StreamLine project (mostly the Sztaki group for ML algorithms and the TU Berlin group for Flink) will result in operators that will be integrated and tested in our workflows according to the design guidelines established during ASAP.

webLyzard pursued active research collaborations throughout the three years of the ASAP project to identify synergies, maximize impact, and exchange results with other research projects. This includes *past projects* such as **uComp** (CHIST_ERA),[36] which developed a human computation framework to blend knowledge extraction and crowdsourcing methods, and **DecarboNet** (FP7),[37] which developed a collective awareness platform for carbon footprint reductions, as well as *ongoing projects* such as **PHEME** (FP7),[38] 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**[39] that detects emerging stories and assess the reliability of user-generated video content (where WLT leads the system integration). In the fourth quarter of 2016, webLyzard has become an *Associate Member* of the **Big Data Value Association (BDVA),**[40] and also presented the ASAP project during the BDVA Summit in Valencia. We aim to actively pursue follow-up opportunities, building on the technological advances and industry contacts of the ASAP project.

[34] www.cactosfp7.eu
[35] pcj.icm.edu.pl/hpdcj
[36] www.ucomp.eu/
[37] www.decarbonet.eu/
[38] www.pheme.eu
[39] www.invid-project.eu/
[40] www.bdva.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:

- **ASAP Source Code Documentation**
  https://github.com/project-asap/ASAP-documentation
- **Intelligent Resource Scheduler (IReS)**
  The open source platform for optimizing, planning and executing complex analytics workflows in multi-engine environments.
  www.github.com/project-asap/IReS-Platform
- **Telecom Analytics**
  Spark Application for peak detection in call detail records
  www.github.com/project-asap/telecom-analytics
- **Platform for Analytics Workflows**
  Unified, 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
- **Spark Tests**
  Set of Apache microbenchmarks to test an alternative scheduling mechanism.
  www.github.com/project-asap/spark-tests
- **Swan Tests**
  Test cases for Swan in clang and intel-cilkplus-runtime
  www.github.com/project-asap/swan_tests

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

1.6 Industry-Oriented Materials

As a result of ongoing industry-oriented activities, exploitable results of ASAP have been 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 aimed 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 went and will go beyond a specific industry and beyond the defined uses cases. Tailored exploitation actions are targeted at **companies** already collaborating with partners of the ASAP consortium (in various domains, not exclusively those covered by the two use cases), **working groups** and **standardization** bodies, and other **stakeholders** with an interest in big data technologies such as policy makers and NGOs.

2.1 FORTH

The CARV laboratory of the Institute of Computer Science at FORTH has expertise in runtime systems, memory management, distributed systems, and languages, and brought this expertise to a focus on the problem of distributed Analytics computations within the ASAP project. With ASAP, FORTH has strengthened its expertise on several aspects of Big Data computations (scheduling, placement, programming models for expressing these computations, as well as compilation and execution of big data queries). The CARV laboratory plans to continue its research on related problems and apply ASAP technology to new Big Data application domains.

We plan to maintain and continue improving on Spark-Nesting, the recursive-query execution engine developed within ASAP that is based on Spark. We have already located several other applications from astrophysics, biology, and natural language processing that will benefit from the extended programming model developed. Moreover, we plan to use the other ASAP modules internally, as we found they facilitate workflow management for domain experts that are not computer scientists and have assisted us in collaborations outside the ASAP project.

We plan to exploit expertise and collaborations developed within ASAP in the future. Specifically, we would like to look into application domains other than bulk/batch analytics computations as we expect that many of the optimizations developed within ASAP will also be applicable in streaming and low-latency/real-time domain applications. We will explore collaborations with the ASAP partners, targeting problems in embedded systems (Big Data from IoT) and financial applications (mixing big data-warehouse with continuous stream processing).

2.2 UNIGE

Working in the ASAP project gave the collaborators from UNIGE the opportunity to realize their innovative vision in terms of research and engineering for a powerful platform for the management workflows on Big Data analytics. UNIGE intends to continue working in extending and improving this platform, as well as test it on real

workflows of new use cases; the latter may come from the domain of bioinformatics and astrophysics, as UNIGE has a close relationship with the Swiss Institute of Bioinformatics (SIB) and the European Organization for Nuclear Research (CERN). Furthermore, UNIGE intends to exploit the acquired research and engineering experience from ASAP, in order to initiate new efforts for the creation of project proposals on Big Data management, to be submitted to the EU and to the Swiss National Science Foundation (SNSF).

2.3 ICCS

The Computing Systems Laboratory of ICCS has a strong expertise in the areas of Big Data and Cloud Computing, particularly focusing on Big Data management, performance aspects of Big Data Analytics and elasticity of Cloud Computing platforms. Within ASAP, the ICCS team exploited this expertise to design and implement the Intelligent Resource Scheduler (IReS), a platform that optimizes, plans and executes complex Big Data analytics workflows in multi-engine environments. More specifically, in the course of the project issues of elastic resource provisioning, management of large volumes of intermediate data, scheduling and execution of tasks in Cloud environments have arisen and prior work of ICCS members in these areas has helped to effectively resolve them.

Moreover, thanks to ASAP, ICCS has expanded its expertise in the areas of (a) performance modelling of distributed runtimes and datastores, (b) management of Big Data analytics workflows throughout their lifetime and (c) multi-engine environments. The lab members have already expanded their research activities to these areas and will continue conducting research in these fields after the end of the project. ICCS has already appointed 4 ASAP-related diploma theses, which have been successfully completed, allowing undergraduate students to familiarize themselves with the related research areas. ICCS plans to maintain and continue improving IReS, extending it to support streaming engines and their special characteristics. We also plan to enhance IReS with more advanced optimization capabilities for special types of data/queries (e.g., Sparql queries over rdf data). The first step towards this direction has already been made with MusQLE[41], a side system that specifically caters for the optimization and execution of SQL queries over multiple engines.

ICCS will exploit the work of ASAP in the following ways: (a) by conducting research in the new fields of expertise gained through ASAP, (b) by collaborating with the project partners in research activities as well as future potential research projects, (c) by internally using parts of the ASAP platform both for our research activities and for managing Big Data related laboratory exercises conducted by undergraduate students

[41] V. Giannakouris, N. Papailiou, D. Tsoumakos and N. Koziris: MuSQLE: Distributed SQL Query Execution Over Multiple Engine Environments. In Proceedings of the 2016 IEEE International Conference on Big Data (BigData 2016), December 5-8, 2016, Washington DC, USA.

as part of the “Distributed Systems” and “Advanced Topics on Databases” courses of the National Technical University of Athens and (d) by exploiting the newly gained expertise in the ongoing projects SELIS[42] and ACTiCLOUD.[43]

2.4 QUB

The High-Performance and Distributed Computing group at QUB has expertise in parallel computing, programming languages and runtime systems. In recent years, it has expanded its core research from the high-performance computing domain into data analytics, cloud computing, and recently fog computing as modern areas where computing at the physical limits of the system (e.g., peak performance, energy constraints) allow application of our core knowledge. Concomitant with this move, the group has become part of a newly created Centre for Data Science and Scalable Computing in the ECIT-2 institute at QUB. The ASAP project, but also related EU projects such as CACTOS, RAPID, UniServer and VINEYARD, have been effective vehicles for QUB to expand its knowledge in the area of data analytics and data centre applications. ASAP in particular, has provided new knowledge to us around the computational and data management challenges in various types of data analytics workloads. It has provided us with a vehicle to further develop our existing work on parallel programming languages and runtime system support, compilation and runtime scheduling algorithms.

In the future, QUB will build on its acquired knowledge to develop a position as a worldwide leader in high-performance analytics. It will feed its acquired knowledge into running projects and also develop new research projects in this area. It will continue development of the Swan programming language using use cases from high-performance computing and data analytics. It will moreover disseminate its experience through undergraduate teaching. In particular, 4 final-year projects investigating the technologies developed in ASAP have been completed and 4 are in progress. QUB will sustain research collaborations with ASAP partners aiming to further develop the ASAP technologies.

2.5 Internet Memory Research

During the project, IMR has reorganized its business activities towards structured data extraction at large scale. Crawling, wrapping and classification have been focused on eMarketPlaces, and the product-related information obtained from the process are currently used as a support for B2B and B2C services. Deliverable 8.4 introduces two such services that have been used during ASAP to evaluate the data processing modules produced by our partners. They are based on complex data analytics

[42] www.selisproject.eu/
[43] acticloud.eu/

workflows whose design and evaluation has been greatly helped by the cooperative work in ASAP. The high-level, platform-agnostic approach adopted by the project has encouraged our big data engineering team to (i) abstract their task at the appropriate level, (ii) view a workflow definition under a modular perspective, and (iii) systematically inspect the behavior of each component, in terms of performance, robustness, and quality. Using these disciplined conception and implementation guidelines has numerous benefits. First, it saves time and efforts when some of the components have to be replaced. It is well known that maintenance and evolution of software is one of the biggest cost in the application development process. We decided for instance to change the classification algorithm which gave poor precision results. With a well-defined workflow made of nearly independent component, the task turned out to be much easier than if we had kept a monolithic conceptual perspective. Second, the ASAP interfaces supply a clear and understandable view of a workflow definition that helps non-expert users to understand what is at stake in a workflow execution, and can be used to exploit various materialisations of the workflow that aims at other goals than the core objective of performance optimization. We used the materialization mechanism for instance to implement a testing version of the classification workflow that can be used to evaluate a new component before introducing it in the catalogue of alternatives.

In the future, we will benefit from the project's lessons to continue the enhancement of our workflows management. One of our next projects is the adoption of Flink as a distributed execution engine, at least for parts of our data processing workflows. This will involve the replacement of some operators with a materialization based on Flink, and their evaluation to check the benefits of the new platform. We plan to generalize the principle of maintaining a catalogue of independent materialization of a given operator, either for choosing the best alternative in a specific evaluation context, or as a possibility to backtrack to a tested and validated implementation if a new one appears to be not satisfying enough. Overall, we will build on the ASAP lessons to improve our technical approach and reduce the cost of our big data processing tools.

2.6 WIND | Tre

WIND contributes to communicating the project results via internal corporate channels as well as through the Web-based channels of the Company's group. Furthermore, WIND exploitation activities are dealing with the impact of the ASAP project in the area of privacy-aware mobility mining to improve our portfolio of services, taking into account the expertise gained. These activities will be enhanced through the use of the ASAP platform, enabling easier large scale data analysis aiming at trend discovery and smart decision making. WIND as a TLC Operator is also interested in Privacy-Aware Mobility Mining to improve its portfolio services according to the legal context: in the new ecosystems the right to the protection of the private sphere must coexist with the right to access to knowledge and to services as a common good. This

evaluation may be useful for different situations like the transportation (things/people) reconstruction and optimization and the Urban/Country Territory promotion for a new smart customized solution.

The results obtained from ASAP will allow WIND to define a roadmap to introduce big data innovations. The project's platform will be useful to analyze different scenarios and realize customized solutions according to both B2B and B2C adopted models. This is in line with the privacy-proven approach that enabled new services using a prototyping approach, thereby reducing initial investments. The experience gained from ASAP has allowed WIND to define a number of possible applications areas:

- Beginning with the Telecommunications Data Analytics Application (TDA) and enriching it with the use of CRM data in order to offer to our Business Market solutions that meet the needs of an emerging market such as in the case of tourism applications with digital solutions that contribute to improving the mobile services offer.
- The capability to better configure the customers offer and differentiate itself from other telco operators (user targeted applications/personalized applications and services).

Using the TDA it would be possible to obtain:

- Forecasts of traffic flows linked to events and/or places (Transportation/Urban Mobility/Tourism).
- Improved customer experience since the traffic forecasts for each event signal the need to improve network capabilities to offer better perceived quality (Transportation/Urban Mobility).
- New "digital assistant" which will be offered to enhance customer interaction.
- A competitive advantage by leveraging Big Data, WIND's 'mine', to expand the range of solutions for an easy "Digital Life".
- An important innovation in the Italian digital market by using the «privacy proved» data.
- Contributions to the development in the Italian and European markets; big data skills and new professions like data scientists.

As an added value, such data can be mined and correlated in order to help business analysts predict patterns and define efficient marketing and business strategies taking into account the specific context of "smart" tourism management. The use cases give WIND the opportunity to create different components for innovative applications: most efficient solutions to offer good services and a good perceived quality of the connections to gain trust from connected customers.

This will be done also by improving and expanding the network capacity in specific areas where a higher user concentration and services usage has been observed as a result of the ASAP framework use.

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 strategic and advanced use of Big Data will help WIND to gain information/insight on the success rate of an event and to improve the effectiveness of WIND released services in the context of the selected event.

For example, as part of WIND exploitation activities in ASAP there was a collaboration with the Mobility Agency of the City of Rome in the area of tourism analysis. 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 ASAP Telecommunication Data Analytics (TDA) application.

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). The TDA would be useful to improve the customer experience according to their potential interests while various planned events in their cities happen. In addition, the TDA may be customized for other unexplored commercial contexts.

Feedback from WIND’s Big Data Department ”Customer Experience and Big Data Analytics” suggested that the use of “thermal” diagrams for the visualization of the activity level of users might be a desirable feature in a series of future applications (similar to the GROOVE visualization reported by webLyzard in Deliverable D6.2). Also concerning other forms of visualization to support some possible applications in WIND there is the so called “O/D (Origin/Destination) Matrix as explained in the TDA application (refer to Deliverable D9.2 and D9.4). Both visualization techniques in future applications are one of the exploitation results of the ASAP platform in WIND’s upcoming services based on Big Data analytics.

Finally, from the point of view of the company, the ASAP activities have allowed 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.

2.7 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 represent highly valuable assets. They are among the primary exploitation targets and essential for the continued growth of webLyzard, and for achieving its long-term commercial goals.

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. Embedded into the ASAP dashboard (D6.4), the visualization components support ad hoc exploration of dynamic datasets (the comparison is not restricted to brands, but can also include other entities such as products, persons and organizations).

Word-of-mouth and active collaboration with large 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) helped to disseminate ASAP results to an international audience. Improved scalability increases the WLT knowledge base, attracts new clients and represents an important competitive advantage, particularly in conjunction with the new visualization methods of WP6. An early exploitation of developed technologies in real-world applications complemented the ongoing evaluation efforts reported in D6.5.

In Year 1 of ASAP, this included the U.S. Climate Resilience Toolkit[44] 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 of T6.1 were integrated into the analytics view of this application, which enables communication experts at the National Oceanic and Atmospheric Administration (NOAA) to monitor Web content streams and continuously improve the toolkit's knowledge repository.

In *Year 2*, our work concentrated on modularizing the core platform and providing a REST API (Application Programming Interface) to upload, annotate, retrieve and visualize structured and unstructured data. The capability to integrate and visualize third-party data will enable joint exploitation together with the other ASAP industry partners, and increase the dissemination potential of international collaborations.

[44] toolkit.climate.gov

In Year 3, we (i) completed the system and interface integration to prepare the public release of the ASAP dashboard (reported in D6.5),[45] (ii) extended our collaboration with the *United Nations Environment Programme* (UNEP), testing new methods developed within ASAP as part of the UNEP Live Platform,[46] and (iii) launched the US Election 2016 Web Monitor as an independent initiative.[47] Such showcases demonstrate technological leadership and increase international visibility, and in turn help attract additional clients who license the developed technologies. The media interest in these showcases has been used not only to promote innovative information services, but also to point towards the ASAP project as a driver of innovation behind a number of enabling technologies.

Looking ahead, the release of a publicly available version of the *ASAP Dashboard* will convey the significant technical progress achieved, supporting both individual and joint exploitation activities:

- *Individual Exploitation* will be pursued both in terms of using the new API framework to attract customers who require semantic and visual search services for their in-house data assets, as well as extending existing platforms such as the *U.S. Climate Resilience Toolkit*. Its comprehensive portfolio of interactive *visualization services* is a core value proposition of webLyzard. The availability of a 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 even a *Container-as-a-Service* (CaaS) approach when deployed in conjunction with the Docker platform.

- *Joint Exploitation***.** The integration and visualization of structured and structured content, as demonstrated through the joint processing of *telecommunications data* (WIND) and *Web content metrics* (webLyzard, IMR), also paves the way for joint exploitation, for example in the form of business intelligence services that relate actual human behaviour (e.g., phone calls and SMS messages during a public event) with peaks of online coverage and the aggregated perceptions of online communities.

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

[45] asap.weblyzard.com
[46] www.uneplive.org/webintelligence
[47] www.weblyzard.com/us-election-2016-web-monitor

Appendix A: ASAP Fact Sheet

ASAP Project – Adaptable Scalable Analytics Platform

www.asap-fp.eu • @ASAP_EU

The **ASAP FP7 Research Project** provides 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 pursued four main goals:

- A general-purpose task-parallel programming model in conjunction with a runtime system for execution in the cloud. The runtime incorporates and advances 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 enables analytics experts to amend submitted tasks in later processing stages.
- A visual analytics dashboard to show query results and metadata in an intuitive manner, with special focus on the interactive exploration of datasets, dynamic temporal controls, on-the-fly query refinement mechanisms, and the geospatial projection of structured and unstructured data (asap.weblyzard.com).

Use Cases and Applications

The generic nature of the ASAP architecture supports a wide range of different tasks. Within the project, the consortium focused 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 extended and enriched 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 extended 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 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 demonstrates 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 statistical indicators. ASAP investigated 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 | Tre
  www.windtre.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 D10.6
Yearly Report for Exploitation and Dissemination of Foreground

WP 10 - Exploitation and Dissemination

Nature: Report
Dissemination: Public