



PROTEUS

Scalable online machine learning for predictive analytics and real-time
interactive visualization

687691

D6.5 Report on scientific dissemination activities – V1

Lead Author: Hamid Bouchachia

Reviewers: Alireza Rezaei Mahdiraji, Marcos Sacristán Cepeda

Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	May 31 th 2017
Actual delivery date:	May 31 th 2017
Version:	3
Total number of pages:	15
Keywords:	Scientific dissemination

Abstract

The present deliverable describes the scientific dissemination activities and materials along with the time line and success indicators. It includes a record of activities related to scientific dissemination that have been undertaken during the first half of the project and those planned for the second period. It reflects on the publications, software and events at which the consortium members took part.

Executive summary

The scientific dissemination has a significant role in raising the impact of the project, contributing in particular to the visibility as well as the awareness about the objective and results of the research undertaken.

In PROTEUS, various scientific dissemination actions have been considered. They are directed to different audiences from the specialised public to specific public in order to illustrate the ambitions of the project and to enhance its global impact. So far PROTEUS has considered many activities that promote the discovery or general understanding of its scope, originality and the results obtained.

Specifically, various types of outcomes have been produced. These outputs have been dedicated to different audiences such as academia, general public, media, policy makers, and the business community.

This report describes the dissemination activities undertaken in the first period of PROTEUS. It mainly presents the activities according to the following classification:

- Dissemination events
- Scientific publications
- Software components

A discussion about these activities is given before a brief outline of future activities is highlighted.

Document Information

IST Project Number	687691	Acronym	PROTEUS
Full Title	Scalable online machine learning for predictive analytics and real-time interactive visualization		
Project URL	http://www.proteus-bigdata.com/		
EU Project Officer	Martina EYDNER		

Deliverable	Number	D6.5	Title	Report on scientific dissemination activities – V1
Work Package	Number	WP6	Title	

Date of Delivery	Contractual	M18	Actual	M18
Status	version 3.0		final <input type="checkbox"/>	
Nature	report <input checked="" type="checkbox"/> demonstrator <input type="checkbox"/> other <input type="checkbox"/>			
Dissemination level	public <input checked="" type="checkbox"/> restricted <input type="checkbox"/>			

Authors (Partner)	BU			
Responsible Author	Name	Hamid Bouchachia	E-mail	abouchachia@bournemouth.ac.uk
	Partner	BU	Phone	+44 1202 96 24 01

Abstract (for dissemination)	The present deliverable describes the scientific dissemination activities and materials along with the time line and success indicators. It includes a record of activities related to scientific dissemination that have been undertaken during the first half of the project and those planned for the second period. It reflects on the publications, software components and events at which the consortium members took part.
Keywords	Scientific dissemination, scientific events, scientific publications.

Version Log			
Issue Date	Rev. No.	Author	Change
March 26 th , 2017	V.0	Hamid Bouchachia	Structure of the document
May 05 th , 2017	V.1	Hamid Bouchachia	First draft
May 18 th , 2017	V.2	Alireza Mahdiraji, Marcos Marcos Sacristán Cepeda	Review comments
May 22 nd , 2017	V.3	Hamid Bouchachia	Final version

Table of Contents

Executive summary 3
Document Information 4
Table of Contents 5
List of Tables 6
1. Introduction 7
2. Dissemination Events and Publications 8
 2.1. Dissemination events 8
 2.2. Publications 9
3. Software 12
4. Forthcoming Actions 14
5. Conclusion 15

List of Tables

Table 1: Dissemination activities 8
Table 2: Events attended 8
Table 3: List of publications 10

1. Introduction

To enhance the visibility of PROTEUS, various dissemination activities have taken place during the first period of the project. In particular, the scientific dissemination under its different forms has been continuous taking a number of guidelines into account:

- Background Intellectual Property
- Foreground Intellectual Property
- Intellectual Property Rights of the partner (patents, copyright, trademark, etc.)
- Confidentiality of data (emanating mostly from industrial partners)
- Target prominent journals and conferences when publishing
- Open access whenever possible
- Rigorous referencing in all publications
- Diversification of the audiences
- Acknowledgment of the EC financial support

This document provides the details of the scientific dissemination activities undertaken in the context of PROTEUS. These include mainly publications, scientific event organization and attendance, as well as software developed during the project. The goal of the document is to highlight the ongoing effort of PROTEUS to present the outcome and enhance the awareness of various communities about the objectives and results of PROTEUS.

The remaining part of this document dedicated to scientific dissemination only consists of 2 sections. In the first of them, the dissemination events and publications that PROTEUS has produced so far are presented. Section 3 describes the software components developed. In Section 3, a brief outline of the next actions to be taken in the future to further strengthen the scientific dissemination effort towards the large public is given.

2. Dissemination Events and Publications

The consortium has been active targeting a number of venues for disseminating the work so far achieved. In particular, various types of outcomes have been produced as shown in Table 1 below. These outputs have been dedicated to different audiences such as academia, general public, media, policy makers, and the business community.

Table 1: Dissemination activities

	Academia	General public	Media	EU policy makers	Business
Scientific Publication	✓	✓			✓
Talk	✓	✓		✓	✓
Poster	✓				✓
Organisation of Conferences	✓				
Social media			✓		

In the following, the details of the dissemination activities are presented.

2.1. Dissemination events

Members of the PROTEUS consortium took part at many international scientific events to disseminate the project objectives, impact and scientific results to a wide audience. The following table summarises the events where PROTEUS was presented.

Table 2: Events attended

Event	Date	Type	Audience	Coverage
The ACM SIGMOD Conference	26/06- 01/07/2016	Scientific	1000	International
Beyond MR Workshop co-located with the ACM SIGMOD Conference	26/06- 01/07/2016	Scientific	50	International
FiCloud Conference	23/08/2016	Scientific	50	International
Predictive APIs Conference	10-12/10/ 2016	Business	500	International
Big Data Spain	17-18/11/ 2016	Scientific / Business	1000	International
EuroPro Workshop in EDBT/ICDT	21-24/03/2017	Academic	200	International
Workshop STREAMEVOLV co-located with the European Conference on Machine Learning	19-23/09/2016	Scientific	400	International
Workshop on "Machine Learning Algorithms and Applications" co-located with the International Conference on Machine Learning and Applications	18-20/12/2016	Scientific	300	International
Big Data Coruña Summit	25-26/05/2017	Scientific / Business	50	International
Big Data Value Association Summit	1-2/12/ 2016	Scientific / industry / policy makers	450	European

In addition to presentations at academic conferences, the members of the consortium have delivered public lectures such as: tutorial (European Conference on Machine Learning), keynote (The IEEE International Conference on Evolving and Adaptive Intelligent Systems), and presentations about PROTEUS in business/academia meetings (BDVA, BD Coruña, BD Spain).

The project members organized international events such as the *Machine Learning Algorithms and Applications* workshop co-located with the *International Conference on Machine Learning and Applications* and the *STREAMEVOLV* workshop co-located with the *European Conference on Machine Learning*.

2.2. Publications

PROTEUS is engaging in academic dissemination of results through high-impact journals and conferences. So top dissemination venues have been targeted like IEEE Transactions on Neural Networks and Learning Systems, Pattern Recognition Letters, the European Conference on Machine Learning, and the ACM SIGMOD International Conference on Management of Data.

The consortium members have submitted research work (under review) in very prominent venues such as Journal of Machine Learning Research, and Pattern Recognition, while some other publications are under preparation for others top journals and conferences in the areas of Machine Learning and Big Data.

It is worthwhile stressing that PROTEUS is primarily targeting journals with high-impact factors: IEEE Transactions on Neural Networks and Learning Systems (IP=4.85), Pattern Recognition Letters (IP=1.58), Journal of Machine Learning Research (IF=2.45), Pattern Recognition (IF=3.39) to illustrate the quality of the work being developed in the context of PROTEUS.

Note also the funding source (EC) has been acknowledged in most of the publications, especially those being under review or are under preparation.

In terms of open access, all partners are aware about the importance of making the outcome of PROTEUS accessible to the general readers. PROTEUS as a consortium as well as the individual partners are primarily pursuing green open access. Since gold open access is not free, PROTEUS partners are asked to seek institutional contribution to cover the open access fees. Presently, most of the publications are made accessible, when they are not gold open access. They are provided in the form of “accepted version for publication” or “published version” after the embargo date (e.g., 6 months from the publication date) to avoid any copyright issues.

Moreover, many of the publications are joint private-public (PP) publications. Some of them are still under review. The consortium will be submitting all publications to EC via SyGMA.

Table 3 summarises the list of publications produced so far. It is very important that submitted publications (under review) and those under preparation are not included.

Table 3: List of publications

No	WP	DOI	Type	Repository Link	Title	Authors	Title of the Journal/Proceedings/Books series	Conf date Journal/Book details	Pages	ISBN/ISSN/eISSN	Publisher	Open access/made available	Peer Rev.	Joint PP pub.	Sub. to EC
1	WP3	10.1145/2926534.2926540	Conf. paper	http://dl.acm.org/citation.cfm?id=2926540	Bridging the gap: towards optimization across linear and relational algebra	A. Kunft, A. Alexandrov, A. Katsifodimos, V. Markl	The 3rd ACM SIGMOD Workshop on Algorithms and Systems for MapReduce and Beyond	26/06/16	-	ISBN: 978-1-4503-4311-4	ACM	YES	YES	YES	YES
2	WP3	10.1145/2882903.2899396	Conf. paper	http://dl.acm.org/citation.cfm?id=2882903.2899396	Emma in Action: Declarative Dataflows for Scalable Data Analysis	A. Alexandrov, A. Salzmann, G. Krastev, A. Katsifodimos, V. Markl	ACM SIGMOD International Conference on Management of Data	26/06/16	2073-2076	ISBN: 978-1-4503-3531-7	ACM	NO	YES	YES	YES
3	WP3	-	Magazine paper	http://sites.computer.org/debull/A15dec/p28.pdf	Apache Flink: Stream and Batch Processing in a Single Engine	P. Carbone, S. Ewen, S. Haridi, A. Katsifodimos, V. Markl, K. Tzoumas	Bulletin of the IEEE Computer Society Technical Committee on Data Engineering	Vol. 38 No. 4, Issue on Next-Generation Stream Processing Systems, 12/ 2015	28-38	-	IEEE	YES	YES	NO	YES
4	WP3	10.1145/2949741.2949754	Journal paper	http://dl.acm.org/citation.cfm?doi=2949741.2949754	Implicit Parallelism through Deep Language Embedding	A. Alexandrov, A. Katsifodimos, G. Krastev, V. Markl	ACM SIGMOD Record	Vol. 45, No .1 03/2016	51-58	ISSN:0163-5808	ACM	YES	YES	NO	YES
5	WP5	10.1109/W-FiCloud.2016.46	Conf. paper	http://ieeexplore.ieee.org/abstract/document/7592720/	An incremental approach for real-time Big Data visual analytics	I. G. Fernández, R. C. Tejedor, A. Bouchachia	The IEEE 4th International Conference on Future Internet of Things and Cloud	22-24/08/16	177-182	ISBN: 978-1-5090-3946-3	IEEE	YES	YES	YES	YES
6	WP5	10.1016/j.patrec.2016.08.015	Journal paper	http://www.sciencedirect.com/science/article/pii/S0167865516302173	Improving the efficiency of IRWLS SVMs using parallel Cholesky factorization	R. D. Morales, Á. N. Vázquez,	Pattern Recognition Letters	Volume 84, 1 December 2016	91-98	0167-8655	ELSEVIER	NO	YES		YES

7	WP4	10.1109/TNNLS.2016.2614393	Journal paper	http://ieeexplore.ieee.org/document/7605500/	A Bi-Criteria Active Learning Algorithm for Dynamic Data Streams	S. Mohamad, A. Bouchachia, M.S. Mouchaweh	IEEE Transactions on Neural Networks and Learning Systems	N/A (early access)	1-13	ISSN:2162-2388	IEEE	YES	YES	NO	YES
8	WP4	N/A	Conf/Work Paper	http://eprints.bournemouth.ac.uk/24798/	Aggregation Algorithm Vs. Average for Time Series Prediction	W. Jamil, Y. Kalnishkan, A. Bouchachia	ECML/PKDD 2016 Workshop on Large-scale Learning from Data Streams in Evolving Environments	STREAMEVOLV-2016, 23 September 2016	69-82	-	Ecole des Mines, France	YES	YES	NO	YES
9	WP4	N/A	Conf/Work Paper	http://eprints.bournemouth.ac.uk/24798/	Active Learning for Data Streams under Concept Drift and concept evolution	S. Mohamad, M.S. Mouchaweh A. Bouchachia	ECML/PKDD 2016 Workshop on Large-scale Learning from Data Streams in Evolving Environments	STREAMEVOLV-2016, 23 September 2016	51-68	-	Ecole des Mines, France	YES	YES	NO	YES
10	WP4	10.1109/EAIS.2015.7368803	Conf. paper	http://ieeexplore.ieee.org/document/7368803/?reload=true&arnumber=7368803	A non-parametric hierarchical clustering model	S. Mohamad, A. Bouchachia, M.S. Mouchaweh	The 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)	01-03/12/2016	1-7	ISBN: 978-1-4673-6698-4	IEEE	YES	YES	NO	YES
11	WP4	10.1142/9789814675017_0005	Book chapter	http://www.worldscientific.com/worldscibooks/10.1142/9548	Fuzzy Classifiers	A. Bouchachia	Handbook on Computational Intelligence	01/05/2016		ISBN: 978-981-4675-00-0	World Scientific	YES	YES	NO	YES
12	WP4	10.1007/s00500-015-1946-4	Journal paper	http://link.springer.com/article/10.1007/s00500-015-1946-4	MSAFIS: an evolving fuzzy inference system.	J. J. Rubio, A. Bouchachia	Softcomputing	Vol. 21, No. 9 May 2017,	2357 - 2366	Online ISBN: 1433-7479	Springer	YES	YES	NO	NO

3. Software

A number of software components have been developed in the context of PROTEUS. In the following a short description of each of them is presented.

a) PROTEUS Engine

This is an overhauled version of Apache Flink supporting hybrid computation on batch datasets and data streams. Details about PROTEUS Engine can be found here: <https://github.com/proteus-h2020/proteus-engine>

b) PROTEUS Language

A declarative language library, called *Emma*, has been developed to meet the need of Scalable Data Analysis. *Emma* aims at improving the developer productivity by hiding parallelism aspects behind a high-level, declarative API. *Emma* supports state-of-the-art dataflow engines like [Apache Flink](#) and [Apache Spark](#) as backend co-processors. More information about the language is available at <http://emma-language.org> and further details are located at: <https://github.com/proteus-h2020/proteus-language>

c) PEACH (Proteus Elastic Cache)

Peach is a distributed key-value cache that can be used both inside and outside of the Proteus scope. The cache aims to provide low latency responses on a distributed elastic deployment with fault-tolerance capabilities. As a generic design, the cache could be integrated within [Apache Flink](#) to speedup computing processes. More information can be found at: <https://github.com/proteus-h2020/peach>

d) PROTEUS Incremental Analytics

A backend module that implements incremental version ($\sim O(1)$ computational cost using approximations) of most common analytics operations. Proteus-backend is implemented on top of the Apache Flink streaming engine. Further details can be found at: <https://github.com/proteus-h2020/proteus-backend>

e) SOLMA (Scalable Online Machine Learning and Data Mining Algorithms)

A scalable library adapted to the data analytics platform, Apache Flink. It consist of efficient distributed online algorithms for basic utilities, sketches as well as advanced online predictive analytics for tasks like classification, clustering, regression, ensemble methods, and novelty and change detection. Currently, the library encompasses a number of algorithms:

- Moments (simple mean, simple variance, weighted mean, weighted variance, exponentially weighted mean and variance, simple moving average, aggregation algorithm).
- Sampling (reservoir sampling, adaptive reservoir sampling, weighted reservoir sampling)
- Heavy hitters (frequent directions algorithm)
- Feature reduction (principal component analysis)
- Event detection (symbolic aggregate approximation)

While some moments, sampling and event detection algorithms are readily available in SOLMA, other are being re-written into SOLMA and will be available very soon. All algorithms are available on Github: <https://github.com/proteus-h2020/SOLMA>

f) PROTEIC.JS

To accommodate visualization and visual analytics for big data, an HTML5 and CSS3 charts library has been developed. It is adapted to work for both data-at-rest and data-in-motion. More details can be found at: <https://github.com/proteus-h2020/proteic>

4. Forthcoming Actions

PROTEUS will continue to develop its scientific dissemination strategy axed on:

- Publications in open access, both gold and green open access
- Publication in prestigious journals and conferences: Currently the consortium team members have submitted a number of publications to high-impact journals and others are under preparation. An edited book on machine learning for big data is under preparation.
- Actions towards the general public, such as the forthcoming festival of learning at BU
- Organisation of dedicated tutorials, special sessions, special issues of journals and workshops such as the STREAMEVOLV workshop in the forthcoming 2017 European Conference on Machine Learning.
- Further development of the software products mentioned in Section 2.3

The website of the project will be updated on a regular basis to ensure that scientific dissemination activities are brought to the attention of both the specialised as well as the general public.

5. Conclusion

The present document provides a short outline of the scientific dissemination activities undertaken in PROTEUS. In particular the list of events the consortium members participated in, a detailed list of publications produced and the list of software components developed in the context of PROTEUS. A brief overview of the future actions is also covered.