PROTEUS

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

687691

D6.1 PROTEUS Business Plan

Lead Author: Giuseppe Maio, Rachel Finn, Graham Hesketh, & Kush Wadhwa, Trilateral Research

With contributions from: Marta Garcia-Barzana(AMIII), Alireza Rezaei Mahdiraji (DFKI)

Reviewer: Hamid Bouchachia (BU)

Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Confidential, only for members of the consortium (including the Commission Services)
Contractual delivery date:	30/11/2018
Actual delivery date:	30/11/2018
Version:	1.0
Total number of pages:	38
Keywords:	Exploitation, Business Plan

Abstract

This deliverable presents PROTEUS partners' exploitation strategy. This revolves around five key exploitation measures, namely using PROTEUS in other R&D projects, leveraging PROTEUS to provide new products and services, using PROTEUS to engage the academic community, using PROTEUS as a teaching tool to develop students' new skills, and exploiting the open-source nature of PROTEUS components. The deliverable presents the already achieved impacts of this strategy and KPIs for future impacts. Notable impacts include: PROTEUS knowledge-transfer to over 60 organisations in Europe, the creation of data-driven impact assessment tools based on partners' PROTEUS-driven recruitment, the engagement of more than 2,000 academics, skill development among 141 students through PROTEUS-based assignment across Europe.

Executive summary

This deliverable describes a five-fold exploitation strategy designed by the PROTEUS consortium to maximise the project impacts following its end in November 2018. The design of this strategy was driven by an EC- based definition of exploitation, according to which, to exploit a project means to carry out at least one of these four actions: A) using project results in further research, B) creating and marketing a new product, C) creating and marketing a new service, D) using the project's results to carry out standardisation activities (IPR Helpdesk 2018). The PROTEUS exploitation strategy was mapped on the EC-based definition above and consists of: using PROTEUS results as a baseline for further innovation in other R&D projects (A); leveraging PROTEUS to provide new products and services (B/C); using PROTEUS to engage the academic community (A); using PROTEUS as a teaching tool to develop students' skills in streaming data analytics (A); exploiting the open-source nature of PROTEUS components (A/B/C/D).

The strategy already produced tangible impacts including:

- the use of PROTEUS-driven technologies in research proposals integrating more than 60 organisations across Europe - this is how the project was exploited in further research (A);
- the employment of PROTEUS-driven data science competencies in the development of data-driven impact assessment tools to combat known risks - this how the project was exploited to create and market new products /services (B/C);
- the addition of online machine learning for streaming data to AMIII's R&D agenda - this is how the project was exploited in further research (A);
- the incorporation of new services for online machine learning in BU and DFKI's commercial offerings and the potential to consult Siemens, SAP, Sondra, Unilever, and Dorset EMC on this topic- this how the project was exploited to potentially create and market new products /services (B/C);
- a great interest of the academic community in PROTEUS-driven research outputs: each publication written by the consortium members was viewed/ downloaded on average over 153 times -this is will maximise the chances of the project results being exploited in further research (A);
- students' skill development thorough engagement with PROTEUS components throughout Europe: in the last three years 141 students in Germany, Spain and the UK completed PROTEUS-based assignments, 2 students carried out their MSc dissertations, and 2 students completed their doctoral degrees on PROTEUS-driven technologies - this is how the project was exploited in further research (A);

The current and proposed KPIs for each of our specific exploitation measures is outlined below. Finally, the document concludes by presenting an example of a potential third party open source business plan aiming to commercialise PROTEUS open source components

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
1.PROTEUS in R&D projects	Number of organisations exposed to PROTEUS innovations	60	100
	Number of funded projects to further develop PROTEUS components	/	3
	Amount of national R&D funding awarded as a result of PROTEUS	£86,000	£500,000
	Companies adopting PROTEUS technologies	/	AMIII
2.Leveraging PROTEUS to provide new products and services	STRIAD (data analytics) customers for Trilateral	/	70
	Consult companies on online streaming data analysis	/	3
3.Using PROTUES to engage the academic community	Additional invitations to collaborate on issues related to hybrid-computation and online machine learning for streaming data	/	5
	Number of downloads of PROTEUS-driven publications	2,000	3,000
4.Using PROTEUS as a teaching tool to develop students' skills	Number of postgraduate students who gained data analytics skills via engagement with PROTEUS components	141	200
	Number of high-level experts on online machine learning / hybrid computation models (PhD holders)	2	6
5. Exploiting the open-source nature of PROTEUS components	Number of organisations contacting partners to implement, enquire about or further develop PROTEUS algorithms or component.	/	20
	Total PROTEUS libraries GitHub commits	15,615	19,000
	Number of organisations contacting partners collaborate or receive support in commercialising PROTEUS components	/	5

Document Information

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

Deliverable	**Number**	6.1	**Title**	PROTEUS Business Plan
Work Package	**Number**	WP6	**Title**	Communication, exploitation and dissemination

Date of Delivery	**Contractual**	M36	**Actual**	M36
Status	version v.1		final ■	
Nature	report ■ demonstrator □ other □			
Dissemination level	Public □ Restricted ■			

Authors (Partner)	Giuseppe Maio, Rachel Finn, Graham Hesketh, Kush Wadhwa (Trilateral Research)			
Responsible Author	**Name**	Giuseppe Maio	**E-mail**	Giuseppe.maio@trilateralresearch.com
	Partner	TRI	**Phone**	+44 (0)207 559 3550

Abstract (for dissemination)	This deliverable presents PROTEUS partners' exploitation strategy. This revolves around five key exploitation measures, namely using PROTEUS in other R&D projects, leveraging PROTEUS to provide new products and services, using PROTEUS to engage the academic community, using PROTEUS as a teaching tool to develop students' new skills, and exploiting the open-source nature of PROTEUS components. The deliverable presents the already achieved impacts of this strategy and KPIs for future impacts. Notable impacts include: PROTEUS knowledge-transfer to over 60 organisations in Europe, the creation of data-driven impact assessment tools based on partners' PROTEUS-driven recruitment, the engagement of more than 2,000 academics, skill development among 141 students through PROTEUS-based assignment across Europe.
Keywords	Business Plan, Exploitation

Version Log			
Issue Date	**Rev. No.**	**Author**	**Change**
05/11/2018	0.1	Giuseppe Maio	
23/11/2018	0.2	Giuseppe Maio	Changes following Rachel Finn's review
29/11/2018	0.3	Giuseppe Maio, Rachel Finn	Changes following Rachel Finn and Kush Wadhwa's reviews
30/11/2018	1.0	Giuseppe Maio, Rachel Finn, Graham Hesketh and Kush Wadhwa	

Table of Contents

List of tables

Abbreviations

BDVA = Big Data Value Association

CASE = Collaborative Awards in Science and Engineering

CROSS- CPP = Ecosystem for Services based on integrated Cross-sectorial Data Streams from multiple Cyber Physical Products and Open Data Sources

EMC = Dorset Engineering and Manufacturing Cluster

ESRC = Economic and Social Research Council

HPC = High Performance Computing

HSM = Hot Strip Mill

I-BiDaaS = Industrial-Driven Big Data as a Self-Service Solution

IF = Individual Fellowships

IP = Intellectual Property

ITN = Innovative Training Networks

KPI = Key Performance Indicators

LASSO = least absolute shrinkage and selection operator

LDA = latent Dirichlet allocation

MSCA = Marie Skłodowska-Curie Actions

ONLMSR = Online Normalised Least Mean Squares regression

R&D = Research and Development

RISE= Research and Innovation Staff Exchange

SAX = Symbolic Aggregate apprXimation

SME = Small and Medium Enterprises

SME = Subject Matter Expertise

SOLMA = Scalable Online Machine Learning Algorithm

TU = Technical University

PROTEUS' Partners:

AMIII = ArcerlorMittal

BU = Bournemouth University

DFKI = Deutsches Forschungszentrum für Künstliche Intelligenz

TREE = Treelogic

TRI = Trilateral Research

1. Introduction: PROTEUS Exploitation Strategy

The purpose of this section is to review the technical documentation on H2020 post-project exploitation and develop an actionable strategy for the exploitation of PROTEUS. This strategy will ensure an impact legacy for PROTEUS following the termination of the project, i.e. November 2018.

This section will review the EC-based definition of exploitation including the actions that it entails. Subsequently, it will present a set of five exploitation measures constituting the Consortium's strategy for the maximisation of PROTEUS impacts past project-end, which will be mapped on the EC-based definition of exploitation. The remainder of this deliverable deals with these five exploitation measures in detail.

According to the European IPR Helpdesk, exploitation is "The utilisation of results in further research activities other than those covered by the action concerned, or in developing, creating and marketing a product or process, or in creating and providing a service, or in standardisation activities" (2018, p.13). Hence, exploitation means carrying out at least one of these four actions: A) using project results in further research, B) creating and marketing a new product, C) creating and marketing a new service, D) using the project's results to carry out standardisation activities. Considering both the exploiting actors and the exploited products, exploitation can be direct or indirect. Exploitation is direct either when partners carry out one or more of the actions above themselves (European IPR Helpdesk 2015, p. 7), or when what is exploited is the direct outcome of the project, i.e., a new software, technology, etc. Exploitation can be indirect either when partners ensure results are used by other entities outside the consortium (European IPR Helpdesk 2015, p. 7), or when it is not the direct outcome of the project to be exploited, but for instance the skills, competences, and tools developed as a result of participation in the project. The exploitation of project results leads to their commercialisation, signifying "turning products and services into a commercially viable value" (European IPR Helpdesk 2016, p.5). The actions above directly or indirectly generate commercial value. Partners have designed PROTEUS exploitation strategy around actions A), B), C), and D) to maintain the project's impact legacy beyond the term of the funded project.

PROTEUS exploitation strategy is made up of five key measures:

1. Using PROTEUS results as a baseline for further innovation in other R&D projects;
2. Leveraging PROTEUS to provide new products or services;
3. Using PROTEUS to engage the academic community;
4. Using PROTEUS as a teaching tool to develop students' skills in online streaming data analytics;
5. Exploiting the open-source nature of PROTEUS components.

The consortium mapped these measures directly to the four actions making up the EC-driven definition of exploitation.

Partners are using project results in further research (A) through: using the results of PROTEUS as a baseline for further innovation in other R&D projects (1), using PROTEUS to engage the academic community (3), using PROTEUS components as teaching tools to develop students' skills in online streaming data analytics (4), and exploiting the open-source nature of PROTEUS components (5).

Partners are creating and marketing new products and services (B and C), through: leveraging PROTEUS to provide new products or services (2) and exploiting the open-source nature of PROTEUS components (5).

Finally, partners are using the project's results to carry out standardisation activities (D) through exploiting the open-source nature of PROTEUS components (5).

This five-fold strategy is a blend of direct and indirect exploitation measures with regards to both the exploiting actors and the exploited products. As far as the exploiting actors are concerned, measures 1 and 2 are examples of direct exploitation: consortium members are the initiators as well as the beneficiaries of these exploitation activities. Measure 3 lies between direct and indirect exploitation: partners have written academic publications, conference papers, white papers, and reports; these outlets are expected to be picked up by and benefit the wider academic community. While partners have used these publications to demonstrate their expertise and support the pursuit of further research in this space, the use of PROTEUS' results by the wider academic community constitutes an indirect exploitation activity. Finally, 4 and 5 are an indirect exploitation measure, as they actively promote external communities' engagement with PROTEUS results: these measures benefit students, data-scientists, system-architects, software developers and steel industry/manufacturing sector professionals.

As far as the exploited products are concerned, partners mainly carried out direct exploitation: the main products exploited are the outcomes of PROTEUS, i.e. SOLMA, LARA and PROTEIC. However, in some cases, i.e. for measure 2, partners carried out indirect exploitation leveraging the skills and competences which stem from the PROTEUS project, but are not a direct product of it. For instance, the skills of the professionals recruited to carry out project tasks are not a direct product of PROTEUS, yet they were exploited in measure 2. This five-fold exploitation strategy is particularly effective as it covers a full spectrum of direct and indirect exploitation measures in respect to both the exploiting actors and the exploited products. The table below provides a summary of PROTEUS partners' exploitation strategy.

The consortium wishes to acknowledge that this deliverable was originally intended as a business plan for an integrated technical PROTEUS product, with a business plan developed by the former partner Lambdoop. However, the way that the PROTEUS project has developed over time has meant that the components of PROTEUS are best exploited more separately by individual partners and for specific purposes. As such, although this deliverable includes a business plan for exploiting the open source components of PROTEUS, it is more correct to regard the deliverable as an exploitation report and plan for PROTEUS components and partners than a business plan for an integrated PROTEUS product. Nevertheless, the final section presents an example of a potential third party open source business plan aiming to commercialise PROTEUS open-source components.

Table 1 Summary of PROTEUS Exploitation Strategy

PROTEUS exploitation measures	Corresponding EC definition	Direct vs Indirect (Exploiting Actor)	Direct vs Indirect (Exploited Product)
1.Using PROTEUS in other R&D projects	A) using project results in further research	Direct: done by partners	Direct: SOLMA/LARA/PROTEIC
2.Leveraging PROTEUS to provide new products and services.	B) creating and marketing a new product / C) creating and marketing a new service	Direct: done by partners	Direct/Indirect: SOLMA/LARA/PROTEIC as well as machine learning skills of those involved in PROTEUS.
3.Using PROTEUS to engage the academic community	A) using project results in further research	Direct/Indirect: done by partners, but picked up by wider academic community	Direct/: SOLMA/LARA/PROTEIC
4.Using PROTEUS components as teaching tools to develop students' skills	A) using project results in further research	Indirect: active promotion of students' engagement with PROTEUS	Direct: SOLMA/LARA/PROTEIC
5.Exploiting the open-source nature of PROTEUS components to maximise its impact	A) using project results in further research / B) creating and marketing a new product / C) creating and marketing a new service / D) using the project's results to carry out standardisation activities	Indirect: community-driven	Direct: SOLMA/LARA/PROTEIC

2. Using PROTEUS results as a baseline for further innovation in other R&D projects

This section delves into partners' exploitation of PROTEUS in R&D activities. The most exploitable PROTEUS-driven technologies in further R&D activities are: SOLMA, an open-source machine learning library specialised in online learning; LARA, a high-level language to express streaming and online machine learning algorithms; and PROTEIC, an open source interactive visualisation library.

BU, DFKI and TRI have applied for European Commission's R&D funding leveraging the technologies developed within PROTEUS in H2020 calls for proposals and plan to continue to do so after the close of the project. Some partners, for example TRI, also intend to leverage PROTEUS to obtain national R&D funding. Hence, national and EU R&D funding bodies are the main targets of this exploitation measures. Finally, partners such as AMIII and DFKI plan to leverage PROTEUS for internal R&D activities: AMIII intends to further investigate the causes of steel defects.

2.1 PROTEUS in H2020 R&D projects

BU has leveraged PROTEUS' experience and technical results in other H2020 R&D projects, i.e., the ICT-11-2018-2019 call on *HPC and Big Data enabled Large-scale Test-beds and Application*. For this proposal, BU emphasised SOLMA's ability to cope with big data challenges such as velocity and volume; these are PROTEUS' key selling points that will provide BU, one of SOLMA's technical developers, with an advantage in big data-related proposals, bids and future R&D projects at the EU level. Starting from the hybrid computation engine for data-at-rest and data-in-motion through which SOLMA runs, BU plans to implement the library in an HPC setting., which is likely to greatly improve the performance of SOLMA's algorithms. Measures of latency, throughput and the velocity of the algorithms are likely to benefit a great deal from the adaption of the SOLMA's abstraction on an HPC setting.

DFKI leveraged PROTEUS in a proposal for the ICT-2-2018 call on *Big Data Technologies and Extreme-Scale Analytics.* PROTEUS ability to process hybrid data schema made up of batch and streaming data renders it particularly relevant for future projects on extreme-scale analytics. DFKI intends to use PROTEUS technologies in any further R&D endeavours at the EU level concerning streaming data algorithms. DFKI were, in fact, the main developers of LARA, a high-level language for streaming online algorithms, which has a great exploitable potential in further R&D projects, given its innovativeness. BU plans to use LARA to create new abstractions similar to SOLMA but tailored to extreme-scale analytics in different business use cases ranging from the manufacturing industry to the agricultural sector.

While BU and DFKI directly leveraged PROTEUS products and new technologies in further R&D, TRI did so indirectly. Through the skills and competences of the staff recruited to partake in PROTEUS, TRI participated in an ICT-11-2018 call where it proposed the implementation of a data cleaning tool, facilitating data set preparation prior to the analysis, through neural network and deep learning techniques. This is an example of indirect exploitation: the skills TRI acquired through PROTEUS-driven recruitment will be employed beyond the research areas circumscribed by PROTEUS to address deep learning- and neural network-type of problems.

Finally, all partners plan to participate in proposals responding to H2020, FP9 & BDVA PPP calls on big data analytics and machine learning: consortium members continually push data analytics forward and partaking in these calls will allow them to further contribute to big data analytics in Europe.

2.1.1 Impacts and KPI for future impacts

As far as the impact of this exploitation measure is concerned, each of these consortia are large and comprise members from across the EU 28. In consequence, at least 60 organisations, including industry partners like Vodafone, Everis, BMW, Philips as well as numerous SMEs and major European mathematical and engineering departments (e.g., Polimi, UPM and others) are now aware of the main innovations of PROTEUS and considering how their existing technological solutions can be integrated with its main components. Should these projects be funded, these European organisations specialised in ICT, software development, machine learning and data science will directly implement PROTEUS components in new platforms, tools and use cases.

R&D activities revolving around streaming data analytics will be likely to increase in number and relevance with time. As a matter of fact, in a 2014 Forrester report on Big Data and Streaming Analytics, Gualtieri and Curran maintain that the streaming application programming models are unfamiliar to software developer (in Freeman 2016): this unfamiliarity will spur further research. Through PROTEUS, partners have placed themselves at the forefront of Europe's R&D streaming analytics research agenda, functioning as catalysts for an EU-wide research-based development of these technologies. This is a very positive impact stemming from the PROTEUS project.

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
1.PROTEUS in R&D projects	Number of organisations exposed to PROTEUS innovations	60	100
	Number of funded projects to further develop PROTEUS components	/	3

As far as the KPI for future impacts, the consortium intends that by 2021, more than 100 organisations will be made aware of PROTEUS innovations and be considering implementing PROTEUS results to support further innovation. In addition, securing funding for at least three H2020 projects intending to leverage PROTEUS results between 2018 and 2021 would be a very positive result, given the nature of the competition to secure H2020 R&D funding.

2.2 PROTEUS in national R&D projects

As mentioned, TRI leveraged PROTEUS indirectly applying for R&D funding through the data science skills of the new personnel recruited as a result of its involvement in PROTEUS. As such, TRI is exploiting their participation in PROTEUS to build new technical tools using the machine learning and data visualisation expertise required for their PROTEUS tasks. As part of this work, TRI applied for internal funding within the UK, i.e. the Defence and Security Accelerator (DASA) and the Innovate UK funding.

2.2.1 Impacts and KPI for future impacts

Trilateral is currently developing a strategy risk assessment methodology for assessing the risk of modern slavery for Ministry of Defence operations. The methodology will include data visualisations using open data on modern slavery to provide users with insights for situational awareness purposes.

The project (£86,000) is funded by the Defence and Security Accelerator (DASA). This represents a strong positive impact for the consortium, showing the positive effect of an indirect exploitation of PROTEUS-enabled skills. Regarding KPI for future impacts, Trilateral aims to secure a further £500,000 in national R&D funding, i.e., Innovate UK funding, to further develop the tool and transform it for use in various sectors in order for it to become Trilateral's first commercial data analytics product.

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
1.PROTEUS in R&D projects	Amount of national R&D funding awarded as a result of PROTEUS	£86,000	£500,000

2.3 PROTEUS in internal R&D projects

Regarding internal R&D activities, DFKI's endeavours will include: the further implementation of new streaming algorithms within SOLMA, and a performance comparison of hybrid computation with other streaming systems supporting batch and streaming data in one engine. In addition, DFKI will further develop the LARA component of PROTEUS, a high-level language for streaming and online machine-learning, allowing it to support not only relational and linear algebra, but also graph data. DFKI will also consider how to apply the LARA language to other PROTEUS components. This internal development will give DFKI an advantage when applying for R&D funding both nationally and at the EU level.

PROTEUS has provided a boost to AMIII's internal R&D activities, laying the foundations for further research on stream analysis algorithms, on one hand, and interactive visualisation dashboards, on the other hand. With regards to the stream analysis algorithms, these represent an area of interest to AMIII due to its applications in predictive maintenance as well as anomaly detection. AMIII's R&D team is committed to implement new applications of streaming data analysis which focus on these two areas. In order to do so, the team will build upon the SOLMA library, in particular the LASSO and ONLMSR algorithms, and take advantage of the huge volume of streaming data that the HSM produces to test these algorithms. SOLMA's algorithms developed within PROTEUS will be used by AMIII's R&D team as benchmarks against which to assess the performance of novel anomaly detection and predictive maintenance algorithms internally developed. Stream analysis algorithms entered, through PROTEUS, AMIII's research agenda and they have become an important part of its R&D endeavours.

As far as the interactive visualisation dashboards are concerned, PROTEIC will be used by AMIII as an initial prototype for further internal development. AMIII's R&D team intends to enhance the current visualisation dashboards developed by TREE to help HSM operators with their daily decisions. AMIII is committed to further develop the PROTEIC components that shed light on the production process and constitute a handy decision-making tool for plant operators with potential to contribute to cut production costs. Furthermore, AMIII's R&D team plans to extend the use of PROTEIC outside the HSM: R&D team members will perform the necessary adjustments that will allow factory operators to use PROTEIC in other AMIII's facilities.

.

2.3.1 Impacts and KPI for future impacts

The addition of online streaming data analysis in AMIII's R&D agenda as a result of PROTEUS is a notable impact for partners and this is likely to have a positive effect on smart manufacturing in Europe, considering AMIII's role as leading steel manufacturer in the EU. As for KPI for future impacts, having PROTEIC installed and used within AMIII's facilities will be an excellent result for the consortium. This adoption is likely to take some time given the conservative nature of the AMIII's management, but its impact will be significant.

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
1.PROTEUS in R&D projects	Companies adopting PROTEUS technologies	/	AMIII

3 Leveraging PROTEUS to provide new products and services

TRI, BU and DFKI have all leveraged PROTEUS to provide new consultancy services focused on data analytics. While BU and DFKI did so directly through the exploitation of PROTEUS technology, TRI did so indirectly through leveraging the skills and competences acquired through PROTEUS-driven personnel recruitment and internal investments.

3.1 Direct exploitation of PROTEUS technologies to provide new products and services

BU and DFKI, the main developers of the PROTEUS technologies, intend to leverage the algorithms, code, and the computational infrastructure resulting from PROTEUS for the provision of new consultancy services. Once again, the most exploitable PROTEUS components are SOLMA, LARA and PROTEIC.

BU delivered a talk at the Dorset Engineering and Manufacturing Cluster (EMC), an UK initiative promoting cross-sector collaborations between UK-based manufactures, scientists and engineers. In this setting, BU informed the regional industrial community about the research activities and the results of PROTEUS. BU plans to provide consultancy services to the Dorset EMC that revolve around online streaming algorithm leveraging the PROTEUS experience.

In addition, BU engaged with a great deal of industry representatives who showed an interest in PROTEUS technologies. Interested parties were from the steel, wood, food, energy, and healthcare sector:

- SIDENOR Aceros Especiales, Spain
- Södra Sweden
- Unilever UK
- Schlumberger, UK
- XIM, UK

More companies will be targeted attending tradeshows, conferences and through leveraging BU's social media channels, including LinkedIn, YouTube and Twitter. BU plans to consult these and carry out PROTEUS-driven knowledge transfer in the industrial sector.

DFKI has had meetings with R&D managers from various industrial sectors who visit DFKI regularly to receive updates on technology and software development. These R&D managers were all introduced to PROTEUS technologies, in particular to the algorithms of the SOLMA library and LARA language, developed by DFKI. Companies who gained exposure to PROTEUS include:

- Siemens
- Samsung
- SAP
- Sony

More companies will be targeted attending tradeshows, conferences and through leveraging DFKI's social media channels, including LinkedIn, YouTube and Twitter. DFKI can consult these companies on online streaming data analysis building on the PROTEUS experience. In particular, DFKI can leverage LARA, one of its main technical contributions, to design software solutions dealing with online machine learning and designed to address companies' specific needs.

3.1.1 Impacts and KPI for future impacts

Table 2 . Summary of Organisation Engaged by PROTEUS Partners

Partner	Organisation/Sectors Engaged	Organisation/Sector Description
BU	SIDENOR Aceros Especiales, Spain	Sidenor is a steel company, leader the production of special steel long products and one of the main producers of forged and cast pieces.
	Södra Sweden	Södra is Sweden's largest forest-owner association, with 51000 forest owners as its members. They process forest to wood products and pulp at their mills.
	Unilever UK	Operating in over 70 markets around the world, Unilever Food Solution's portfolio include Knorr, PG Tips, Flora, Colman's and Hellmann's.
	Schlumberger, UK	Schlumberger is the world's leading provider of technology for reservoir characterization, drilling, production, and processing to the oil and gas industry.
	XIM, UK	Xim is a creative, entrepreneurial SME focussing on innovative Digital Health applications delivering R&D projects across Europe.
DFKI	Siemens	Largest manufacturing company in Europe active in several divisions such as industry, energy and healthcare.
	Samsung	Multinational conglomerate with 174 billion USD in 2016.
	SAP	German. Base European multinational software corporation; they specialise in enterprise software for business operations and customer relations. Its 2017 yearly revenue was 23.46 EUR
	Sony	Japanese multinational specialising in customer and professional electronics,

		gaming, entertainment and financial services

The companies above were all introduced to BU and DFKI's new data analytics services driven by PROTEUS technologies and revolving around online streaming data analysis. In the case of DFKI, the companies they interacted with are multi-nationals such as Siemens, SAP, Samsung and Sony. BU interacted with both multinationals and SMEs, i.e. Unilever, XIM, and Sondra. Siemens operates in the smart manufacturing sector where PROTEUS technologies are most readily applicable. This is a good impact brought about by the PROTEUS consortium as it may lead to an industry-wide domino effect in the adoption of PROTEUS-driven technologies. BU played a crucial role in presenting PROTEUS in various sectors outside the steel manufacturing, including food, healthcare and wood industries. Regarding the KPI for future impacts, providing commercial consultancy services for three private companies on online streaming data analysis will be an excellent result for the PROTEUS consortium.

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
2.Leveraging PROTEUS to provide new products and services	Commercial consultancy clients on online streaming data analysis	/	3

3.2 Indirect exploitation of PROTEUS to provide new products and services

TRI used the expertise acquired through PROTEUS-driven recruitment to build a technical team which has helped expand its current service portfolio and used PROTEUS to demonstrate expertise in machine learning that has supported these endeavours.

TRI's involvement in the PROTEUS project allowed the company to expand its research consultancy offerings and add data analytics services and products to its portfolio. In order to work on PROTEUS, TRI built a multi-disciplinary technical team of data scientists with expertise in maths, physics, engineering, statistics and computer science. The investment made possible by PROTEUS funding will allow the team to continue working within TRI following PROTEUS project's end in Novemebr 2018.

In the final stage of the PROTEUS project, TRI's technical team began internal R&D exercises to plan three potential data analytics products and services. These products and services include:

- Air Pollution App
  - Using a number of open-datasets, this app would predict air pollution using noise pollution data gathered through smart-phones.
- STRIAD
  - This product will be a cloud-based data-driven risk management solution which strengthens users' abilities to address vulnerability achieving effective and efficient evidence-based decision-making, within the policing, community safety and public sector in general. TRI is currently applying for Innovate UK funding to support the development of this tool.
- Modern Slavery Risk Assessment Tool
  - Development of a strategy risk assessment methodology for assessing the risk of modern slavery for Ministry of Defence operations. The methodology will be

integrated into STRIAD for piloting and will include data visualisations using open data on modern slavery to provide users with insights for situational awareness purposes. The project is funded by the Defence and Security Accelerator (DASA).

TRI intends to commercialise these products and services to:

- Law Enforcement Agencies - Lincolnshire Police
- Community Safety Partners - Peterborough Safety Partnership
- Regulatory Bodies - Gangmasters Licensing Authority
- Security Practitioners - NWG Network
- Central UK Government - Stabilisation Unit, Ministry of Defence

3.2.1 Current impacts and KPI for future impacts

TRI will consider the exploitation of PROTEUS as a success if they are able to commercially deploy STRIAD to at least five organisations in each category by the end of 2019. This commercial revenue stream within Trilateral will be a direct result of their participation in PROTEUS and the technical team recruited because of TRI's participation in the project.

Table 3. Number of target client KPIs in each sector in the next 3 years

Partner	Sectors to be engaged	Target clients per sector per year		
TRI		2019	2020	2021
	Law Enforcement Agencies	5	8	10
	Community Safety Partners	5	8	10
	Local Councils	5	8	10
	Regulatory Bodies	5	8	10
	Enforcement Agencies	5	8	10
	Security Practitioners	5	8	10
	Central UK Government	5	8	10

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
2.Leveraging PROTEUS to provide new products and services	STRIAD (data analytics) customers for Trilateral	/	70

4 Using PROTEUS to engage the academic community

Partners have made significant efforts to engage the academic community through conference and tradeshows attendance, publication of articles in peer-reviewed journals, and presentation of conference papers.

The table below shows a selection of the events attended by the PROTEUS consortium where partners engaged the academic and scientific communities.

Table 4: Scientific and academic events attended by the PROTEUS partners

Event	Date	Type	Audience	Coverage
RP1				
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
Workshop STREAMEVOLV co-located with the European Conference on Machine Learning	19-23/09/2016	Scientific	400	International
Big Data Spain	17-18/11/ 2016	Scientific / Business	1000	International
Big Data Value Association Summit	1-2/12/ 2016	Scientific / industry / policy makers	450	European
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
EuroPro Workshop in EDBT/ICDT	21-24/03/2017	Academic	200	International
Big Data Coruña Summit	25-26/05/2017	Scientific / Business	50	International
RP2				
European Big Data Standards Workshop, Dublin 2017	14 August 2017	Scientific / industry / policy makers	50	European
Flink Forward Berlin 2017	12 September 2017	Scientific / industry	100	International
The Growing Ubiquity of Algorithms in Society, The Royal Society, London 2017	13 September 2017	Scientific / industry / policy makers	50	UK
Workshop at the European Big Data Value Forum: Industrial Data Platforms for the Manufacturing domain	21st to 23rd November 2017	Scientific / industry / policy makers	500	European
LibreCon 2017, the Business and Open Technologies Conference	October 19-20, 2017	Scientific / industry	300	European

AI Congress 2018, London	30-31 January 2018	Industry / Scientific	1000	International
SIRIUS Digital Twins workshop, Oslo	20 March 2018	Scientific / industry	30	European
18th UK Workshop on Computational Intelligence	September 5-7, 2018	Scientific	80	International
Workshop on Software Architecture Challenges in Big Data	25 September 2018	Scientific / industry	25	European
Smart Factories Workshop	26 September 2018	Scientific / industry	80	International
European Big Data Value Forum, 2018 Data driven business model workshop	November 13-15 2018	Scientific / industry	500	European

Below we present a sample list of publications written by PROTEUS consortium members highlighting the number of downloads or views of each publications as provided by the publishers (publications for which no data was found were excluded from the sample).

Publication	**# Downloads /views**
Bridging the gap: towards optimization across linear and relational algebra by A. Kunft, A. Alexandrov, A. Katsifodimos, V. Markl.	*210*
Emma in Action: Declarative Dataflows for Scalable Data Analysis by A. Alexandrov, A. Salzmann, G. Krastev, A. Katsifodimos , V. Markl.	*219*
Implicit Parallelism through Deep Language Embedding by A. Alexandrov, A. Katsifodimos, G. Krastev, V. Markl.	*117*
An incremental approach for real-time Big Data visual analytics by I. G. Fernández, R. C. Tejedor, A. Bouchachia.	*299 (views)*
A Bi-Criteria Active Learning Algorithm for Dynamic Data Streams by S. Mohamad, A. Bouchachia, M.S. Mouchaweh.	*493 (views)*
Improving the efficiency of IRWLS SVMs using parallel Cholesky factorization by W R. D. Morales, Á. N. Vázquez.	*6 (views)*
Aggregation Algorithm Vs. Average for Time Series Prediction by W. Jamil, Y. Kalnishkan, A. Bouchachia.	*32*
A non-parametric hierarchical clustering model by S. Mohamad, A. Bouchachia, M.S. Mouchaweh.	*77 (views)*
MSAFIS: an evolving fuzzy inference system by J. J. Rubio, A. Bouchachia.	*334*
Benchmarking Distributed Stream Data Processing Engines by S Jeyhun Karimov; Tilmann Rabl ; Asterios Katsifodimos ; Roman Samarev ; Henri Heiskanen ; Volker Markl.	*5 (views)*
Batch-based active learning: Application to social media data for crisis management. by D. Pohl, A. Bouchachia, H. Hellwagner.	*173 (views)*
Active learning for classifying data streams with unknown number of classes by S. Mohamad, M.S. Mouchaweh ,A. Bouchachia.	*24 (views)*
Scalable online learning for Flink: SOLMA library by W. Jamil, N.-C. Duong, W. Wang, C. Mansouri, S. Mohamad, A Bouchachia.	*12*

4.1 Impacts and KPI for future impacts

As shown by the table 4 above, PROTEUS was introduced to a total of 6,215 academics and scientists. On average, each event where PROTEUS was presented was attended by over 310 academics and scientists. This impressive result will maximise the likelihood of consortium members to partake in future research and collaborations.

The titles above cover a wide range of topics including software engineering, data science and statistics. Based on a bibliometric analysis carried out, on average a PROTEUS-related academic publication was viewed/downloaded by over 153 researchers; the PROTUES-related publications above were viewed/downloaded by a total of 2,001 researchers. This shows that PROTEUS' technologies and innovations were well received by the wider academic community who engaged with PROTEUS-driven research outputs. These impacts are likely to lead to further exploitation as they are crucial to publicise PROTEUS results.

This successful dissemination supports exploitation measures as engagements with the academic community are likely to result in future collaborations around under-researched topics such as hybrid-computation models and online machine learning for streaming data. The consortium members agree that five additional invitations to collaborate on issues related to hybrid-computation and online machine learning for streaming data, i.e. scientific publications, knowledge sharing, further research endeavours, would function as a benchmark for success in relation to this activity. Additionally, partners agreed that increasing the total number of downloads of PROTEUS-driven publications to 3,000 by 2021 would be a good benchmark for success.

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
3.Using PROTUES to engage the academic community	Additional invitations to collaborate on issues related to hybrid-computation and online machine learning for streaming data	/	5
	Number of downloads of PROTEUS-driven publications	2,000	3,000

5 Using PROTEUS as a teaching tool to develop students' online streaming data analytics skills

BU, DFKI and AMIII are using PROTEUS components as teaching tools, focussing on the skill development of postgraduate students. Furthermore, partners intend to continue developing students' skills using PROTEUS technologies in industrial PhD studentships, supervised by BU and DFKI, that will include a work placement/ research assignment within AMIII's industrial setting. These studentships will allow further research and innovation on streaming data analysis using PROTEUS components as a baseline and enhance students' skill set and employability. Studentships will be awarded through the BU-led CASE PhD Studentships in the UK, and the DFKI-led Marie Skłodowska-Curie Actions PhD Studentships in Germany.

5.1 PROTEUS as a teaching tool for European postgraduate students

BU used SOLMA as teaching tool for 60 Master students enrolled in the Applied Data Analytics programme who took the courses "Big Data & Cloud Computing" and "Web Mining & Analytics". Presentations on SOLMA were conveyed to these students to provide them with inputs for their MSc dissertation projects on online machine learning. Furthermore, MSc coursework within Applied Data Analytics often built upon SOLMA concepts such as online classification, online clustering and online event detection. BU-driven students' exposure to PROTEUS had a clear impact on their selection of MSc projects: one of these, which has recently been completed, focused on drift detection, while another one, still ongoing, deals with distributed machine learning. As far as PhD theses are concerned, some of the BU PhD students built upon PROTEUS: a PhD thesis was centred around active learning for data streams and was successfully defended in October 2017; another doctoral thesis dealt with online regressions and is expected to be defended in September 2019.

Furthermore, as part of BU's strategy for sustaining PROTEUS research activities, BU has developed two versions of the SOLMA library: a formal Flink-based library and a prototyped version in Matlab, Python and R. The availability of SOLMA in multiple programming languages will greatly benefit its future development which BU will drive through PhD theses, and MSc theses and assignments within the Master's in Applied Data Analytics.

Finally, in addition to the existing "Stream Analytics" MSc course unit that builds on PROTEUS, BU plans to propose another unit within the MSc in Applied Data Analytics called "Smart Systems" that covers some of the algorithmic aspects of PROTEUS.

Regarding DFKI, 75 Computer Science and Informatics MSc students from TU Berlin implemented and evaluated new algorithms within the SOLMA library as part of their Big Data Analytics Project, under the supervision of DFKI academic staff. This is an implementation-intensive 9 ECT project module requiring a substantive amount of time spent on algorithmic design. DFKI plans to propose an undergraduate-level course focussing on implementing distributed streaming algorithms related to SOLMA as well as potential dissertation or bachelor theses based on SOLMA.

PROTEUS has strengthened and intensified the already existing research collaborations between AMIII and the University of Oviedo. Each year, two postgraduate students conduct research for their dissertations under the supervision of AMIII's R&D team, and/or take part in internship programmes within AMIII. Students are from a variety of master's degree courses, including *Condensed Matter Physics & Nanotechnology MSc*, *Mathematical Modelling, Statistics and Computing MSc*, *Telecommunications Engineering MSc*, and *Industrial Engineering MSc*.

During the PROTEUS project, AMIII's R&D team supervised a student's thesis which focussed on streaming data analysis. This project thesis aimed to implement a streaming regression model capable

of adjusting to the changing conditions of the plant (i.e. the changing process parameters, speed of the rolling phase, roller conditions, equipment deterioration, etc.). The project thesis used the PROTEIC dashboard as a visualisation tool to display the regression coefficients; additionally, the project sought to use PROTEIC to observe the extent to which the changing conditions of the plant, as shown by the regression coefficients displayed on PROTEIC, were related to the triggering of the alarms in the plant.

Building on this positive PROTEUS-driven experience, AMIII's R&D team is committed to ensure that one of the two MSc theses they supervise yearly will focus on the analysis of data streams through 2021. In particular, the next MSc project thesis will focus on the PROTEIC visualisation dashboard following up from the previous PROTEUS-based MSc project supervised by AMIII's R&D team.

Table 5. PROTEUS Partners' University Affiliation

Partner	University Affiliation	University Description
BU	Bournemouth University	BU is a partner of the PROTEUS consortium. It offers an undergraduate and a post-graduate degree in data science, the *BSc in Data Science and Analytics* and the *MSc in Applied Data Analytics* respectively. One of the course-units within the MSc, namely *Analytics for Data Streams*, is tightly linked to PROTEUS research activities.
DFKI	Technical University of Berlin	One of the most prestigious research universities in Europe, the Technical University of Berlin is home to DFKI, one of PROTEUS partners. DFKI staff either work within or in close collaboration with TU Berlin.
TRI	University of Manchester	Many of TRI's staff are alumni of the University of Manchester. This is one of the leading research universities in the UK and one of the most important in Europe. A new *MSc in Data Science* has been recently established within the Department of Social Statistics of the School of Social Sciences.
AMIII	University of Oviedo	AMIII has been a partner of the University of Oviedo for many years. The University of Oviedo is a hub for innovation in Asturias, where AMIII's R&D department operates. AMIII and the University of Oviedo initiated in 2014 the *ArcelorMittal Chair for Research and Development* through which 150 research partnerships worth over 4.5 million Euros were signed.

5.1.1 Impacts and KPI for future Impacts

The number of students enrolled in TU's faculty of Electrical Engineering and Computer Science in 2018 was 6,291. Of these, 25 students took the Big Data Analytics Project used SOLMA (via DFKI) and contributed new algorithms within the library as part of their curriculum. Hence, in the last three years at least 75 students in German used PROTEUS innovations directly to improve their data analytics expertise. This is a very significant impact considering post-graduate students are likely to either enter the employment market or continue their studies at PhD level: both destinations constitute ideal settings to further develop their PROTEUS-driven knowledge of online machine learning for streaming data. In addition, Germany has one of the strongest industrial sectors in Europe and having

educated over 75 students using cutting-edge online machine learning algorithm may contribute to closing the skill-gap in the smart manufacturing sector. As a matter of fact, the manufacturing sector now employs more computer and data scientists than factory operators (BSA 2018), in spite of the latter outnumbering the former.

BU already trained 2 high-level experts in machine learning and streaming analytics: one completed her PhD, the other is close to completion. Furthermore, BU, through its PROTEUS-based teaching activities, trained two MSc students in drift detection and distributed machine learning through dissertation projects: they will soon enter the job market more employable, thanks to PROTUES. This is a positive impact for the Department of Computing and Informatics already commended by Times Higher education for students' high level of employability. These graduates will have access the European employment market and academia, given BU's international reputation. The university is home to over 2,600 international students from over 123 countries. Additionally, 96% of BU's research enjoys an international reputation. Students who gained exposure to PROTEUS components were mainly MSc in Data Analytics students, around 20 per academic year. Hence, PROTEUS contributed to train 60 students in 3 years on topic such as online machine learning and advanced data analytics in the UK. This is a crucial impact driven by the PROTEUS consortium, when considering the ties between BU and Dorset EMC. BU students trained through PROTEUS are now more likely to be employed in the manufacturing sector in Britain, contributing to the development of smart manufacturing in the country.

AMIII strengthened its already existing collaboration with university of Oviedo. AMIII introduced PROTEUS to 6 students in 3 years. AMIII and the University of Oviedo signed a number of research partnerships: 2014 the two parties gave birth to the *ArcelorMittal Chair for Research and Development* through which 150 research partnerships worth over 4.5 million Euros were signed. PROTEUS may lead to further research partnerships between AMII and the University of Oviedo around online streaming data analysis, spurring R&D-driven economic growth in Southern Europe.

Regarding KPIs for future impact evaluation, increasing the number of students who gained exposure to PROTEUS-driven technologies will be considered a positive result.

Table 6. Number of students introduced and to be introduced to PROTEUS by 2021

	Students already introduced 2018	Students to be introduced through 2021	Intended total
BU	60	100	160
DFKI	75	100	175
AMIII	6	8	14
TOTAL	**141**	**208**	**349**

For instance, an increase from 75 to 100, from 60 to 100 and from 6 to 8 in the number of students introduced to PROTEUS by 2021 by DFKI, BU and AMIII respectively will be considered an excellent result.

5.2 PROTEUS CASE PhD Studentship

PROTEUS partners intend to co-write 3 CASE PhD studentships that will build upon PROTEUS technologies and will equip students with marketable skills in online machine learning for data streams.

CASE studentships, formerly known as "Collaborative Awards in Science and Engineering", are UK-based research partnerships which allow MA/MSc holders to undertake a research experience leading to a PhD within a mutually beneficial collaboration between academic and industrial partner organisations. During their CASE studentships, students benefit from a great deal of research training opportunities and they receive industrial research experience as well as business training, including project management, logistics, business strategy, finance, etc. Industrial research experience is acquired by students through placements at the headquarter(s) of the industrial partner(s). The placements have a minimum length of 3 months and a maximum length of 18 months. The time to be spent on placement can be accumulated through several shorter placements, if considered appropriate by the partner organisations.

CASE studentships can be awarded in the following ways:

A) Converting Doctoral Training Partnerships' (DTP) Grants into CASE studentship: DTP Training Grants can be considered CASE studentships if students meet the minimum CASE studentship criteria, i.e. carrying out a 3-18-month placement;
B) Through funding of the Economic and Social Research Council (ESRC).

PhD proposals forming the basis for a PhD CASE studentship are co-authored by the academic partner organisations in conjunction with the industrial partner organisations and advertised on the websites of the former. To apply for CASE studentships, students send their CVs and covering letters to the academic and industrial partners who will jointly consider each application.

BU suggested the following title for one of the PROTEUS CASE PhD studentships:

Advanced scalable online clustering algorithms.

This proposal will be included in a bid for funding and sent to the ESRC.

The time-table of a paper-based PROTEUS CASE PhD studentship may be as follows:

Year 1 Paper 1 supervised by BU (2019-2020)

Year 2 Paper 2 supervised by BU and DFKI/ TU Berlin, with one semester to be spent in Berlin (2020-2021).

Year 3 12- month Industrial placement at AMIII, in Asturias, Spain (2021-2022)

Year 4 Paper 3 supervised by BU in collaboration with AMIII (2022-2023).

Supervisors: Professor Hamid Bouchachia (BU); Dr. Alireza Rezaei Mahdiraji (DFKI).

BU is committed to write in conjunction with DFKI 3 CASE PhD proposals in total, whose scope are the development of new online clustering algorithms, in particular online spectral clustering thus enriching SOLMA, like the example provided above.

5.3 PROTEUS European Industrial Doctorate PhD Studentship funded by MSCA-ITN

As part of PROTEUS' strategy to continue research activities based on project's results, partners' will jointly write 3 research proposals in a bid for Marie Skłodowska-Curie Actions funding. This funding will be used for industrial PhD studentships, supervised by DFKI with co-supervision from BU, which include students' research secondments at AMIII's facilities in Asturias, Spain. The Marie Skłodowska-Curie Actions (MSCA) are EC-funded grants aimed at researchers at all stages,

including doctoral candidates and experienced researchers; their purpose is encouraging transnational, intersectoral and inter-disciplinary mobility within the EU and beyond.

There are three main types of MSCA: The Innovative Training Networks (ITN), the Individual Fellowships (IF), and the Research and Innovation Staff Exchange (RISE). ITNs provide support for joint research training and doctoral programmes resulting from the co-operation of universities, research institutes and business/industrial partners within Europe. IF supports experienced researchers promoting their mobility within and outside Europe through grants covering two-year salary, a mobility allowance and research costs. RISE provides support to short-term mobility of research and innovation staff at all career stages, including postgraduate students, senior management staff, administrative and technical staff. ITNs are particularly relevant to PROTEUS: partners plan to write three proposals securing MSCA ITN funding between 2019 and 2020.

ITN's funding for joint research training can, in turn, take the form of Collaborative European Training Networks, European Industrial Doctorates and European Joint Doctorates. In order to boost researchers' employability through ITNs, researchers spend some time conducting research assignments in non-academic/ industrial contexts. To enhance researchers' employability a Career Development Plan is designed jointly by the research and her/his supervisor as part of the ITNs.

PROTEUS partners will jointly write 3 research proposals seeking to secure MSCA funding for ITNs in the form of European Industrial Doctorates. Subsequently, DFKI will publicise PROTEUS European Industrial Doctorate Fellowships on its website.

DFKI suggested the following title for one of the PROTEUS European Industrial Doctorate fellowship funded by MSCA- ITN:

Hybrid computation engine by architectural design

PROTEUS is currently built on top of Apache-Flink streaming API which was then adapted to accommodate batch data: PROTEUS was thus not built as a hybrid technology by design. The idea behind this PhD is building a streaming system which has a hybrid component by design and is able to deal with both streaming and batch data through its architecture. This will enhance the hybrid computation component of PROTEUS and improve the performance of the SOLMA algorithms as well as data process performances of PROTEUS.

The time-table of a paper-based PROTEUS European Industrial Doctorate PhD studentship funded by the MSCA-ITN may be as follows:

Year 1 Paper 1 supervised by DFKI (2019-2020)

Year 2 Paper 2 supervised by DFKI and BU, with one semester to be spent in Bournemouth (2020-2021).

Year 3 12- month Industrial placement at AMIII, in Asturias, Spain (2021-2022)

Year 4 Paper 3 supervised by DFKI in collaboration with AMIII (2022-2023).

Supervisors: Dr. Alireza Rezaei Mahdiraji (DFKI); Professor Hamid Bouchachia (BU).

DFKI is committed to write, in conjunction with BU, 3 PhD proposals in total whose scope is to design a new engine which supports hybrid computation by architectural design.

5.3.1 Impacts and KPI for future Impacts

The PROTEUS industrial PhD programmes, i.e. CASE and MSCA PhD studentships, will train 6 high-level experts in machine learning and streaming analytics in three countries in Europe, i.e. German, Spain and the UK, from 2019 until 2023. These professionals/ academics will actively contribute to shape the EU big data analytics research agenda, focusing on online machine learning for streaming data. This is likely to have a great impact on smart manufacturing as these analytics' technique have a clear industrial use.

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
4.Using PROTEUS as a teaching tool to develop students' skills	Number of postgraduate students who gained data analytics skills via engagement with PROTEUS components	141	208
	Number of high-level experts on online machine learning / hybrid computation models (PhD holders)	2	6

6 Business strategy for PROTEUS open source components

This section highlights partners' commitment to support the exploitation of the open-source nature of PROTEUS components for the benefit of the wider academic, industrial, and software developer communities. It also underlines some of the positive impacts of the open-source release of PROTEUS' components and results. Finally, it presents an example of a potential third party open source business plan aiming to commercialise PROTEUS open-source components. This will show how third parties willing to commercialise PROTEUS could potentially extract value from its open source technologies, as well as how partners can facilitate this third party- driven commercialisation of PROTEUS.

6.1 Open-Source nature of PROTEUS components

Partners ensured that all PROTEUS' deliverables and technical components, source code, visualisation tools, hybrid engine architectural elements, are available as open source. The technical components of PROTEUS are stored on GitHub and freely available for everyone to use:

- SOLMA: https://github.com/proteus-h2020/proteus-solma ;
- Other PROTEUS components: https://github.com/proteus-h2020 .

The publicly available deliverables are accessible through the PROTEUS project website:

- https://www.proteus-bigdata.com/results/public-deliverables/ .

In addition to this, BU's translation of SOLMA from a Flink-based library into a R-, Python- and Matlab- package has greatly contributed to introduce PROTEUS' outputs to the software developer and data science communities in Europe and beyond. This will facilitate an open source community-driven development of SOLMA which will potentially boost its commercial value and thus represents an indirect exploitation measure. Furthermore, thanks to their translation in multiple programming languages, PROTEUS open source algorithmic components could be used as benchmarks for future developments based on the current technologies: AMIII plans to use the KPI measurements of PROTEUS algorithmic components as provided in D6.2 as benchmarks for future PROTEUS-based anomaly detection algorithmic development.

The open-source nature of PROTEUS components, however, will not only be beneficial to the software developer and data-science communities. The open-source nature of the PROTEUS visualisation tool, PROTEIC, developed by TREE in conjunction with AMII, is likely to positively impact the steel industry and more generally the manufacturing sector.

AMIII worked very closely with TREE advising on the variables and the visual components to include in this visualisation dashboard. For instance, following AMIII's recommendations, TREE included in PROTEIC a heating map, an alarm triggering system, and a window signalling errors. The final version of PROTEIC is thus the result of continuous interactions between TREE and AMIII and reflects the industrial expertise of the latter.

PROTEIC thus brings together statistical/ AI techniques for anomaly detection, i.e., the algorithms of the SOLMA library whose results are displayed in PROTEIC and AMIII's subject matter expertise. Using subject matter expertise and AI in the design of an open-source visualisation tool is a novelty in the smart manufacturing sector, often characterised by a lack of full integration of statistical/AI expertise and industrial know-how (Moyen and Iskander 2017). PROTEUS, through its open-source

ethos, will thus contribute to advance knowledge within the sector, providing a freely available template, PROTEIC, which other manufacturers can deploy in various industries.

As a matter of fact, because of the way in which PROTEIC gathers coil data and displays coil variables using unique IDs, any manufacturer producing coils can potentially reuse this dashboard and tap into PROTUES knowledge base. PROTEIC thus provides the basis for further research on innovative ways to integrate AI expertise and industrial know-how in the smart manufacturing sector. This could be greatly beneficial to European Small and Medium-sized Enterprises (SME) working in the manufacturing sector. These accounted for 58% of total employment and 42 % total added-value in the sector in 2016 (EC 2017, p.15).

In order to enable the exploitation of these results, partners will target enterprises in twitter campaigns, tradeshows and academic conference attendance as well as through publications in academic journals and blog posting. The trade-shows to be targeted will include:

- The *Manufacturing and Supply Chain Conference and Exhibition* in Milton Keynes, UK, on the 25th of June 2019;
- The *European Big Data Value Forum* in Helsinki, Finland, in November 2019;
- *Data Works Summit* in Barcelona, Spain, 18-21 March 2019;
- *Data 2019*, Prague, Czech Republic, 26-28 July 2019.

In addition, partners will engage with these communities through posting on social media such as Twitter, LinkedIn and YouTube, as well as through writing posts for blogs such as Data Science Central, IBM's Global Data Science Forum, FastML, datascience@berkeley, FlowingData, Open Knowledge International, Smart Data Collective, Smart Industry Blog, Research- Data Alliance, CODATA.

Table 7 shows a summary of beneficiaries of the freely available PROTEUS GitHub libraries, technical components and deliverables. As shown in the table, partners will continuously market PROTEUS through a range of different channels.

Table 7. Key Beneficiaries of PROTEUS GitHub libraries and Deliverables

Beneficiaries' Category	Beneficiaries	Marketing Channels
Technical Beneficiaries		
Data scientists	Beneficiaries include: independent data scientists, R developers, DataCamp, IBM, etc.	Twitter, Tradeshows, Academic Conferences, Academic Papers.
Software Developers	Beneficiaries include: Izertis, Intermark, Shinetech Europe, etc.	Newsletter, Twitter, Tradeshows.
IT Specialists	Benefiaries include: Keedio, D4RK0studio, Strategy Big Data, Datio Big Data, etc.	Twitter, Tradeshows.
Manufacturing Sector Beneficiaries		
Steel and Manufacturing technical personnel	Steel and Manufacturing Sector SMEs. Large industries should also be targeted. These include: Vale, USIMINAS, Tata Steel, etc.	Tradeshows
Academia		

Academics/ Students	Crucial for stimulating further research based on PROTEUS' algorithms and services. These include: Bournemouth University's Masters in Applied Data Analytics faculty members/students; University of Manchester's Social Research Methods and Statistics faculty members/students; KU Leuven's Master of Artificial Intelligence faculty members/students.	Academic conferences, academic paper collaborations, academic papers.

6.2 Impacts and KPI for future Impacts

PROTEUS partners were successful in engaging small and medium enterprises as well as large industries by attending each year the European Big Data Value forum. PROTEUS was presented in Valencia in 2016, in Versailles in 2017, and in Vienna in 2018. Furthermore, DFKI and BU's engagements with industry representatives from SAP, Sony, Samsung, Siemens, Unilever, and Sidenor successfully brokered the innovation of PROTEUS to software developer market leaders as well as companies in various sectors such as food, wood, and healthcare. BU was successful in engaging small-, medium-, and large- manufacturing enterprises at the Dorset Engineering and Manufacturing Cluster. These represent significant impacts for the PROTEUS exploitation strategy. By 2021, the consortium hopes that more than 20 organisations contacted through the channels above will directly approach consortium members to enquire about PROTEUS open source components or further develop them with support from project partners.

We used the number of PROTEUS GitHub' commits to measure engagement with PROTUES open source components. We found that PROTEUS' open source components have over 15,615 commits on GitHub. Commits in this case indicate the number of individual revisions of PROTEUS software components stored on GitHub. On average, each of the PROTEUS libraries stored on GitHub was modified 821 times. The most popular PROTEUS library was *proteus-engine,* with 12,611 total commits. Although these commits are indicate partners 'revisions of PROTEUS components, they are a good measure of current and future impact as an increase in their number in the next 3 years will show community engagement with PROTEUS, past project end.

Over the next three years, partners hope that our further dissemination of PROTEUS results will result in at least 10 organisations contacting partners to implement, enquire about or further develop PROTEUS algorithms or components. An additional positive impact would be 19,000 total GitHub commits by 2021.

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
5. Exploiting the open-source nature of PROTEUS components	Number of organisations contacting partners to implement, enquire about or further develop PROTEUS algorithms or component.	/	20
	Total PROTEUS GitHub commits	15,615	19,000

6.3 Example of third party's commercialisation of PROTEUS open- source components

Below we present an example of a third party's open source business plan aiming to commercialise PROTEUS open-source components. This will show how third parties willing to commercialise PROTEUS can extract value from its open source technologies to demonstrate PROTEUS' commercial potential. Additionally, it will show how partners can facilitate this third-party driven commercialisation of PROTEUS.

The plan below updates the business plan presented in the last project periodic review. The plan is the result of data-driven business model workshop PROTEUS partners attended at the European Big Data Value Forum in Vienna on the 15th of November 2018. Specifically, the plan has been validated by expert participants at the workshop, including representatives from Siemens, TU Wien (Vienna Technical University), the Lisbon Council, and the Foundation Next Society.

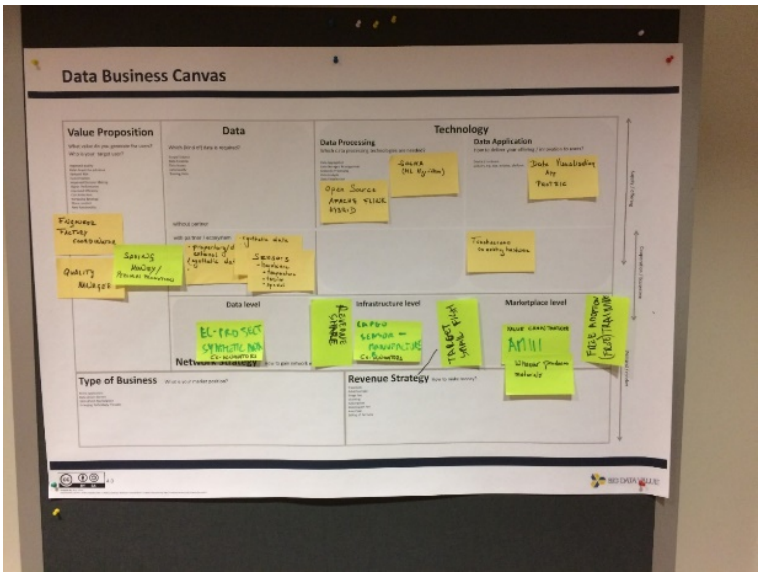

Figure 1: PROTEUS data-driven business model canvas developed at the Data-Driven Business Model Workshop at EBDVF 2018, Vienna

Third party's value proposition

The target user of PROTEUS technologies are manufacturers producing coils, i.e. steel, naval, aero-space industries, etc. However, PROTEUS technologies, in particular the online machine learning tools for streaming data, could be extended to every production line of items made up of a single component, i.e., steel, wood, glass, whose quality is monitored continuously through sensors. The potential target users of the PROTEUS-based products/services developed third parties willing to commercialise its open-source components would be factory managers/operators and quality control managers. The former monitor product quality in the short run, while the latter do so in the long run.

The main job the target user needs to do is predicting defective products before they reach their final stage. This has a great potential for cost savings and allows the re-manufacturing of the defective product. The existing solutions assume no defects and rely on the success rate of the machineries.

While these have a high success rate, failures resulting in the production of defective products are very costly, especially in the steel industry.

Hence, third party could propose to use the PROTEUS open-source code already available on GitHub to predict defects in production before products reach their final stage. Third party's value proposition would consist of a two-level individualisation service, through which the online machine learning algorithms stored in SOLMA could be tailored to specific products, i.e., glass, plastic, etc., as well as specific companies' needs, i.e., defects could be due to temperature, pressure and other issues.

Third party's data

As PROTEUS algorithms were developed using AMIII's proprietary data, third parties commercialising PROTEUS could use synthetic data to market the new PROTEUS-based services they would implement. In particular, third parties could partner with existing projects aiming to develop synthetic data. For example, the I-BiDaaS (2018) project is an EC-funded project which aims to create a data fabrication platform, meaning a platform for generating user-modelled realistic synthetic data for testing and development. One of the domains of I-BiDaaS (2018) is precisely the manufacturing sector.

Third party's Unique Selling Point: PROTEUS Technology

Third parties' commercialisation of PROTEUS open-source components would rely upon the already existing PROTEUS technologies. The data process engine they would use would be the hybrid computation model built on top of Apache-Flink and developed by DFKI. The algorithms to be used would be those available in the SOLMA library, i.e. SAX, LASSO, LDA, ONLMSR, developed by BU and DFKI, as well as new ones which could be developed and stored in the library. The data visualisation tool to be used would be PROTEIC developed by TREE. PROTEIC could be potentially incorporated by third parties willing to commercialise PROTEUS into a user-friendly touch-screen application to be used on smartphones or tablets.

Network Strategy: PROTEUS partners as facilitators of third parties' commercialisation of PROTEUS

Partners' role in this example of a third party's PROTEUS-based open-source business plan would consist of consulting third parties willing to commercialise PROTEUS. Specifically, BU could provide consultations on SOLMA, and DFKI on the hybrid computation model. Additionally, partners could potentially introduce third parties performing the commercialisation of PROTEUS open-source components to both co-innovation partners as well as value chain partners. The former would operate at both the data and the infrastructure level. For instance, the I-BiDaaS (2018) project would act as co-innovator in its role as interim data provider, while sensor producers would act as co-innovators in their role as infrastructure providers. The latter would operate at the market place level and would function as a broker of the new PROTEUS-driven based technologies into the market. Partners could introduce third-parties willing to commercialise PROTEUS to AMIII. Partners are already acting as facilitators of third parties' exploitation of the open-source components of PROTEUS. On the 3rd of December partners will hold a virtual conference with researchers from the EC-funded CROSS-CPP project (2018) to introduce them SOLMA's functionalities.

Third party's IP management strategy

PROTEUS-based new services and products which could be potentially developed by third parties could be licensed commercially: third parties would retain all the revenue deriving from the commercialisation of their PROTEUS-based services and products. For instance, if third parties decided to commercialise PROTEUS open-source algorithmic components by creating a new

algorithm that fits the purposes and needs of a specific company, i.e. their potential customer, they could decide to retain this IP and monetise it.

Market Size

The size of the European Data market is growing exponentially. The value of the data economy was worth almost 300 billion EUR in 2016; this is projected to reach 739 billion EUR by 2020 (EC 2018). Potential targets would be SMEs and other commercial organisations working in various sectors, including smart manufacturing and AI. These would be targeted through blogposts to be published in the following online outlets: Data Science Central, IBM's Global Data Science Forum, FastML, datascience@berkeley, FlowingData, Open Knowledge International, Smart Data Collective, Smart Industry Blog, Research- Data Alliance, CODATA.

KPI for future impacts

The consortium members agree that five invitations to collaborate on the commercialisation of the hybrid-computation and/or online machine learning open-source components of PROTEUS would function as a benchmark for success in relation to this activity.

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
5. Exploiting the open-source nature of PROTEUS components	Number of organisations contacting partners collaborate or receive support in commercialising PROTEUS components	/	5

7 Conclusion

This deliverable described a five-fold exploitation strategy designed by the PROTEUS consortium to maximise the project impacts following its end in November 2018. This strategy consisted of: using PROTEUS results as a baseline for further innovation in other R&D projects; leveraging PROTEUS to provide new products and services; using PROTEUS to engage the academic community; using PROTEUS as a teaching tool to develop students' online streaming analytics skills; exploiting the open-source nature of PROTEUS components.

The strategy already produced tangible impacts including:

- the use of PROTEUS-driven technologies in research proposals integrating more than 60 organisations across Europe - this is how the project was exploited in further research (A);
- the employment of PROTEUS-driven data science competencies in the development of data-driven impact assessment tools to combat known risks - this how the project was exploited to create and market new products /services (B/C);
- the addition of online machine learning for streaming data to AMIII's R&D agenda - this is how the project was exploited in further research (A);
- the incorporation of new services for online machine learning in BU and DFKI's commercial offerings and the potential to consult Siemens, SAP, Sondra, Unilever, and Dorset EMC on this topic- this how the project was exploited to potentially create and market new products /services (B/C);
- a great interest of the academic community in PROTEUS-driven research outputs: each publication written by the consortium members was viewed/ downloaded on average over 153 times -this is will maximise the chances of the project results being exploited in further research (A);
- students' skill development thorough engagement with PROTEUS components throughout Europe: in the last three years 141 students in Germany, Spain and the UK completed PROTEUS-based assignments, 2 students carried out their MSc dissertations, and 2 students completed their doctoral degrees on PROTEUS-driven technologies - this is how the project was exploited in further research (A);

The current and proposed KPIs for each of our specific exploitation measures is outlined below.

Table 8. Summary of exploitation impacts and KPI for future impacts

PROTEUS Exploitation Measures	KPI	Target achieved 2018	Target 2021
1.PROTEUS in R&D projects	Number of organisations exposed to PROTEUS innovations	60	100
	Number of funded projects to further develop PROTEUS components	/	3
	Amount of national R&D funding awarded as a result of PROTEUS	£86,000	£500,000
	Companies adopting PROTEUS technologies	/	AMIII
2.Leveraging PROTEUS to provide new products and services	STRIAD (data analytics) customers for Trilateral	/	70
	Consult companies on online streaming data analysis	/	3
3.Using PROTUES to engage the academic community	Additional invitations to collaborate on issues related to hybrid-computation and online machine learning for streaming data	/	5
	Number of downloads of PROTEUS-driven publications	2,000	3,000
4.Using PROTEUS as a teaching tool to develop students' skills	Number of postgraduate students who gained data analytics skills via engagement with PROTEUS components	141	200
	Number of high-level experts on online machine learning / hybrid computation models (PhD holders)	2	6
5. Exploiting the open-source nature of PROTEUS components	Number of organisations contacting partners to implement, enquire about or further develop PROTEUS algorithms or component.	/	20
	Total PROTEUS GitHub commits	15,615	19,000
	Number of organisations contacting partners collaborate or receive support in commercialising PROTEUS components	/	5

Acknowledgements

We are grateful to Sonja Zillner, organiser of the BDVA's Data-Driven Business Model workshop in Vienna on the 15th of November 2018, Allan Hanbury and Clara-Michaela Dvorak for their valuable contribution and insights in designing an example of a third party's business plan based on PROTEUS open-source components.

References

BSA (2018). *Every Sector Is a Software Sector: Manufacturing How Software Is Turbocharging Manufacturing Opportunities for All*. [online] Software.org. Available at: https://software.org/wp-content/uploads/Every_Sector_Software_Manufacturing.pdf [Accessed 23 Nov. 2018].

Cross-cpp.eu. (2018). *CROSS-CPP*. [online] Available at: https://cross-cpp.eu/ [Accessed 28 Nov. 2018].

European Commission (2017). *ANNUAL REPORT ON SMES 2016/2017: Focus on Self-Employement.* [online] Available at: http://file:///C:/Users/User/Downloads/Annual%20Report%20-%20EU%20SMEs%202016-2017.pdf [Accessed 29 Nov. 2018].

European Commission (2018). *Building a European data economy*. [online] Digital Single Market. Available at: https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy [Accessed 28 Nov. 2018].

European IPR Helpdesk (2016). *Your Guide to IP Commercialisation*. [online] Iprhelpdesk.eu. Available at: https://www.iprhelpdesk.eu/sites/default/files/documents/EU-IPR-Guide-Commercialisation.pdf [Accessed 23 Nov. 2018].

European IPR Helpdesk (2015). *The Plan for the Exploitation and Dissemination of Results in Horizon 2020.* [online] Iprhelpdesk.eu. Available at: https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/FS-Plan-for-the-exploitation-and-dissemination-of-results_1.pdf [Accessed 23 Nov. 2018].

European IPR Helpdesk (2015). *How to manage IP in Horizon 2020: project implementation and conclusion*. [online] Iprhelpdesk.eu. Available at: https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Fact-Sheet-IP-Management-H2020-Project-Implementation-and-Conclusion.pdf [Accessed 23 Nov. 2018].

European IPR Helpdesk (2015). *Fact Sheet Commercialising Intellectual Property: knowledge transfer tools.* [online] Iprhelpdesk.eu. Available at: https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Fact-Sheet-Commercialising-IP-Knowledge-Transfer-Tools.pdf [Accessed 23 Nov. 2018].

European IPR Helpdesk (2018). *Dissemination and Exploitation in Horizon 2020*. [online] Ec.europa.eu. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/other/events/2017-03-01/8_result-dissemination-exploitation.pdf [Accessed 23 Nov. 2018].

Freeman, H. (2016). *Streaming Analytics 101: The What, Why, and How* - DATAVERSITY. [online] DATAVERSITY. Available at: http://www.dataversity.net/streaming-analytics-101/ [Accessed 23 Nov. 2018].

IBiDaaS (2018). *Project's Innovations | IBiDaaS*. [online] Ibidaas.eu. Available at: https://www.ibidaas.eu/Projects-Innovations [Accessed 23 Nov. 2018].

Moyne, J. and Iskandar, J. (2017). "Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing". *Processes*, 5(4), p.39.