This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777533.

PROviding Computing solutions for ExaScale ChallengeS

D1.3	Final Report		
Project:	PROCESS H2020 – 777533	**Start / Duration:**	01 November 2017 36 Months
Dissemination[1]:	PU	**Nature[2]:**	**R**
Due Date:	31 October 2020	**Work Package:**	**WP 1**
Filename[3]	PROCESS_D1.3_FinalReport_v1.0.docx		

ABSTRACT

In this report, the final findings of the project are summarized. The report covers the final status of the five PROCESS use cases (including the validation of the use case requirements), an updated data management plan with respect to the usage of data, and the final evaluation and dissemination of the PROCESS architecture. The report also includes the innovation potential report based on the use case developments. And finally, an overview of the dissemination and engagement plan and market research report.

1 PU = Public; CO = Confidential, only for members of the Consortium (including the EC services).

2 R = Report; R+O = Report plus Other. Note: all "O" deliverables must be accompanied by a deliverable report.

3 eg DX.Y_name to the deliverable_v0xx. v1 corresponds to the final release submitted to the EC.

Deliverable Contributors:	Name	Organisation	Role / Title
Deliverable Leader[4]	Schmidt, Jan	LMU	Coordinator
Contributing Authors[5]	Maassen, Jason; Madougou, Souley	NLESC	Writers
	Cushing, Reggie	UvA	Writers
	Graziani, Mara; Müller, Henning	HES-SO	Writers
	Schmidt, Jan; Höb, Maximilian	LMU	Writers
	Meizner, Jan	AGH	Writer
	Somoskői, Balázs; Pancake-Steeg, Jörg	LSY	Writer
	Bubak, Martin	UISAV	Writer
Reviewer(s)[6]	Habala, Ondrej	UISAV	Reviewer
	Belloum, Adam	UvA	Reviewer
Final review and approval	Höb, Maximilian	LMU	Coordinator

Document History

Release	Date	Reasons for Change	Status[7]	Distribution
0.0	23.07.2020	Initial chapters	Draft	Working Group
0.1	08.09.2020	Draft sections	Draft	Working Group
0.2	23.09.2020	Section-wise drafts	Draft	Working Group
0.3	05.10.2020	Updated full draft	Draft	Working Group
0.4	20.10.2020	Internal review	In Review	Consortium
0.5	25.10.2020	Final internal review	In Review	Consortium
1.0	31.10.2020	Release	Released	Public

[4] Person from the lead beneficiary that is responsible for the deliverable.

[5] Person(s) from contributing partners for the deliverable.

[6] Typically, person(s) with appropriate expertise to assess the deliverable quality.

7 Status = “Draft”; “In Review”; “Released”.

Table of Contents

Executive Summary

This deliverable marks the end of the funding period of the PROCESS project and therefore, gives a brief summary of the six building bricks of the project:

- Pilot Prototypes and Use Cases
- Data Management
- PROCESS Architecture
- Validation
- Innovation
- Dissemination and Engagement Plan and Market Research

In this report, we do not repeat all findings, results, publications and innovation already reported in previous deliverables. We focus on brief summaries and added references to the relevant deliverables. Throughout the project, we included up-to-date information in our reports. In the second period of the project, we focused on finalising the open source software releases including a best practise guide to achieve reusability and sustainability of the outcomes of the project, which is reported in detail in D8.4.

In D1.3 we give a final overview of the five prototypes, our use cases, which all will continue their development after the project, while using the PROCESS ecosystem where ever possible. Especially their needs were analysed in our data management report, which proves the initial Data Management Plan (DMP) as valid.

The core of the PROCESS ecosystem is the architecture, which displays all service modules, the PROCESS building bricks, and their connections. This architecture was finalised and fully described in D4.5 and has been since then presented in several events, which are detailed in this report in Section 3.1.2.

During the lifetime of PROCESS, we validated and evaluated each single service and each single module, but moreover, the interaction between all modules. We built each part of the ecosystem aware of scalability and usability towards Exascale-ready systems. The used mechanism and results are also summarized in D1.3.

From the beginning, all our use cases aimed to generate an impact within their scientific and industrial communities. The resulting innovations can be applied to other domains and included in other ecosystems. Those findings are included in this report.

Finally, we summarise the dissemination and engagement plan (DEP) and market research activities and report on the potential users within the different tier systems.

List of Figures

List of Tables

1 Progress Report of the Use Cases

This chapter gives an overview of the progress of all Use Cases since the last report and their final state. The goal of this section is to highlight the developments since the last reports in the respective deliverables, each Use Case also gives a short overview of the Use Case itself and the most important milestones throughout the PROCESS project.

1.1 Use Case 1

The main outcome of Use Case 1(UC1) is the development of a three-layer software architecture, i.e. CamNet, for training different deep neural network models. The architecture functionalities are shown in Figure 1. Figure 2 presents more details on the three layers. Such an application is modular, easy-to-deploy with Docker containers and offers a wide range of functionalities. Multiple workflows are provided in such a way that they can be used independently to obtain intermediate outcomes, or assembled to reproduce a full pipeline. Each workflow can be shared, reused and updated independently by different institutions. The developed libraries are now available as GitHub repositories [8] [9] [10], and Docker containers [11].

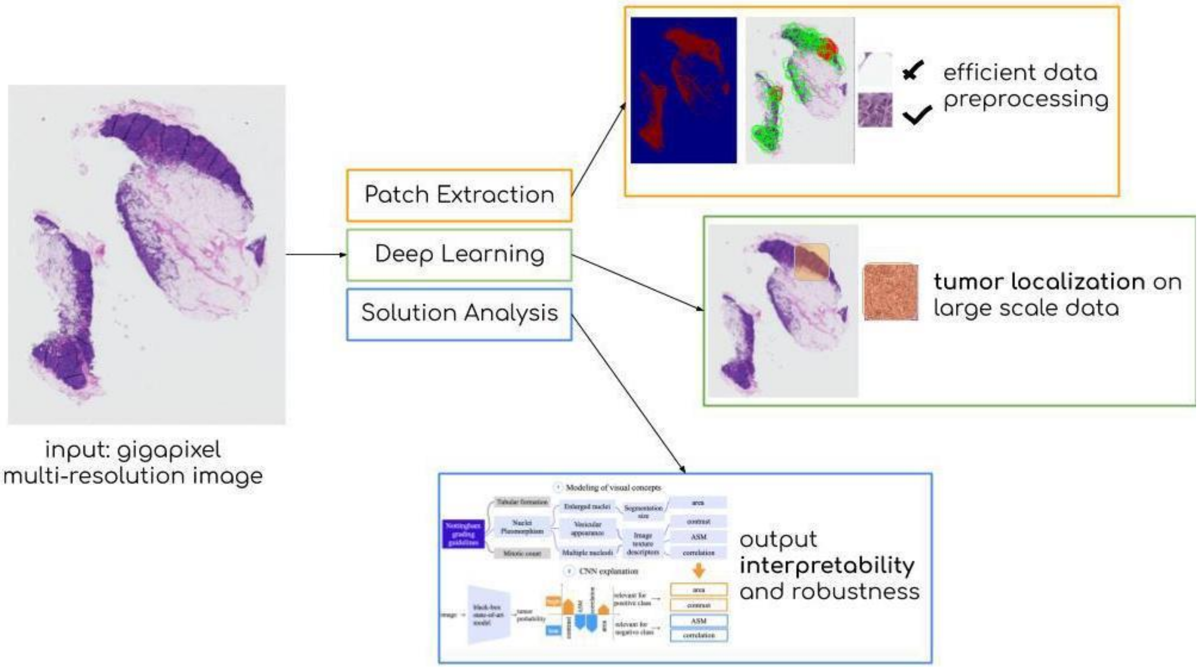

Figure 1: Overview of the CamNet functionalities

[8] https://github.com/medgift/PROCESS_L1

[9] https://github.com/medgift/PROCESS_L2

[10] https://github.com/medgift/PROCESS_L3

[11] https://github.com/ieggel/process-uc1-integration

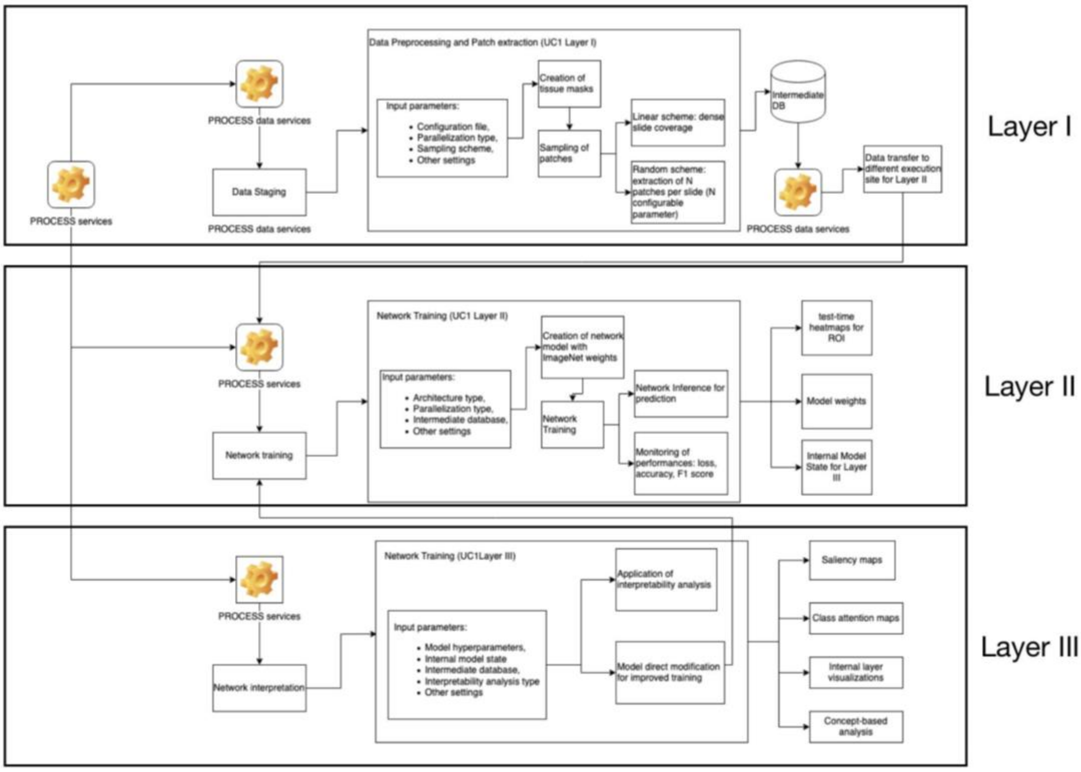

Figure 2: Detailed overview of the CamNet architecture and integration with PROCESS components

Layer I focuses on the patch extraction and data pre-processing of the Whole Slide Images (WSIs) (see Figure 1 in D4.5 and Table 3 in D4.5). The layer scalability and performance were evaluated in terms of execution time against different sizes of the input data, the number of CPU cores and the number of patients as unique WSIs (see Tables 2 and 4 in D8.1, Fig. 1 in D8.1). The time for data transfers between HPC centres was evaluated in D8.1 (see Table 6) and further tested in D3.3 (see Fig. 10 and 11).

No major changes were implemented with respect to D4.5 for Layer I. The use case software was updated in Layer II and Layer III, for which we report the latest development and results in Figure 3.

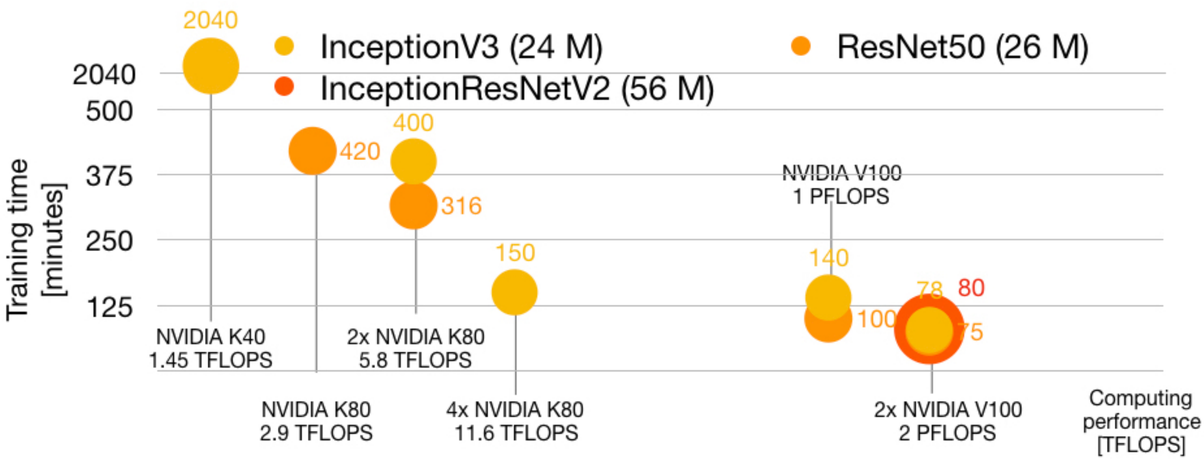

Figure 3: Comparison of average training times of ResNet and Inception models

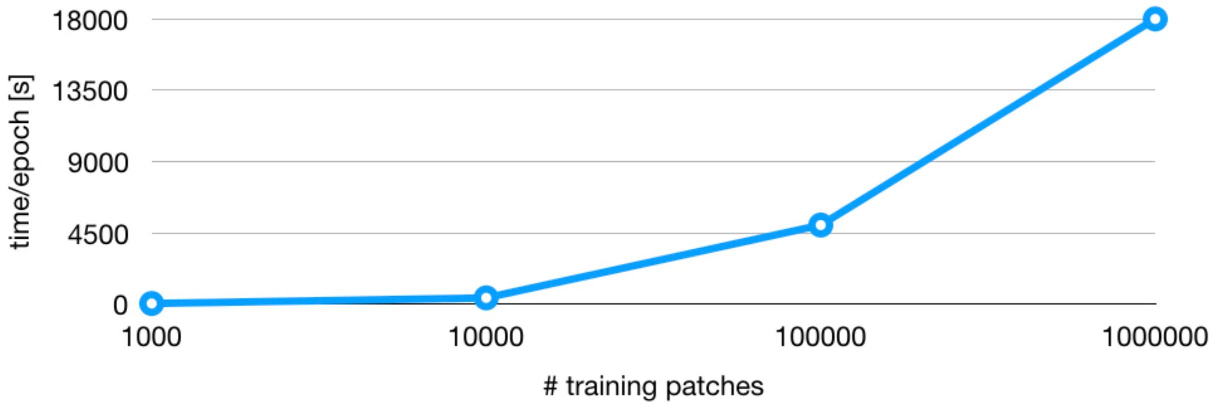

Figure 4: Training time per epoch vs. increasingly larger data sizes for ResNet50 on 2xNVIDIA V100

The support of Horovod and openMPI is now available for Layer II of the application. The code is downloadable either as a GitHub repository or as a Docker container with preinstalled support of Horovod, GPU-functionalities and openMPI[12].

Different parallelization strategies were tested on two testing sites, namely HES-SO and UISAV. We compare in Figure 3 multiple parallelisation schemes. We use 50 Gbs of training data. The number of parameters being trained is reported in brackets (M = millions). The floating-point operations per second TFLOPs of each scheme is reported against the training time. The size of the circle is proportional to the number of parameters in each network

The training times of ResNet50 and InceptionV3 are compared over 10 epochs for 50 Gb of training data**.** As expected, slower GPUs require a longer time to perform the training operations, with a NVIDIA K80 requiring more than 7 hours to train the 26 millions of parameters of ResNet50. This time is reduced by 4 hours and a half when using the latest NVIDIA V100. The model parallelization on two GPUs, particularly on 2 NVIDIA V100 shows the scalability of the network training over multiple GPUs, is completed in less than two hours to train the 24 millions of parameters of Inception V3. The scalability of network training over larger datasets is reported in Figure 4, for ResNet50 distributed on 2 NVIDIA V100. This is indicative of the scalability of the combination of Layer I and Layer II estimating the training time per epoch for increasing dataset sizes.

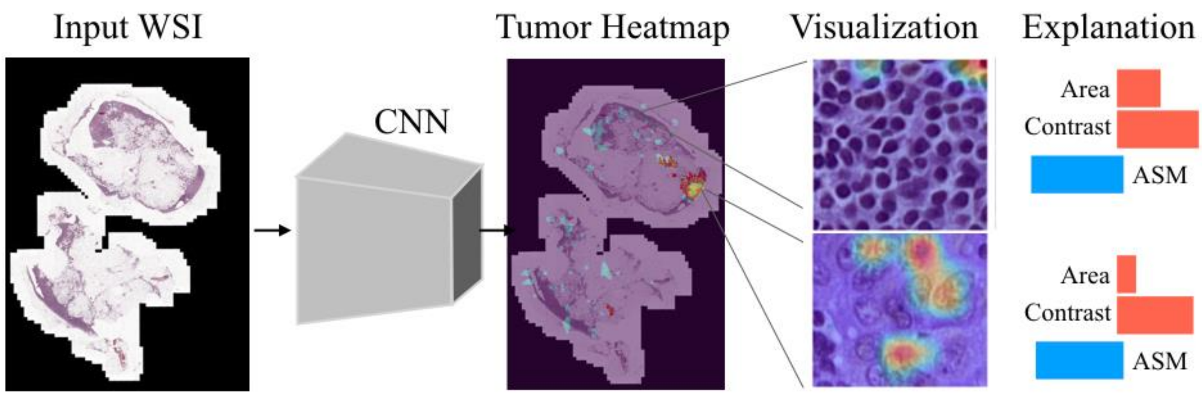

Figure 5: An example of the functionalities in Layer III

Figure 5 shows an example of the functionalities in Layer III, which performs distributed inference for building the tumour heat map and provides interpretability analyses and insights about network training.

---

12 https://github.com/medgift/PROCESS_L2

Layer III is now complete and available on GitHub. The intermediate visualizations in Fig. 1 of D2.2 were improved by performing distributed inference with single- and multi-GPU support. As shown in Figure 6, heat maps are overlaid over the original input WSIs and compared to the manual tumour segmentations provided by pathologists. The inference of nearly 10,000 patches is distributed over 5 processes on a single NVIDIA V100, requiring below 4 minutes to compute the heat map for an input WSI (230 s).

Besides this, the network output is interpreted at the patch-level using gradCAM[13], gradCAM++ [14]and LIME (see the grey boxes in Figure 6. The heat maps of tumour probability are computed by distributed inference over multiple model replicas with single- or multi-GPU support.).

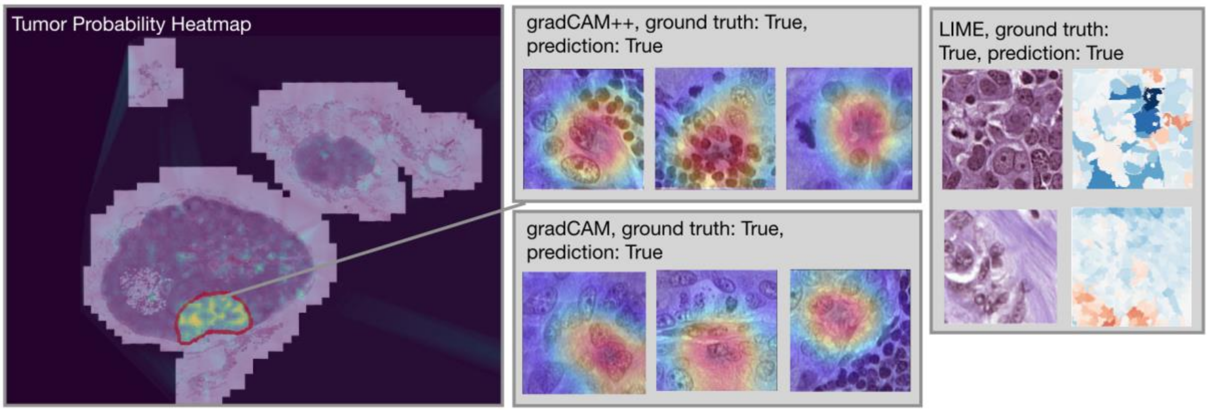

Figure 6: Visualization outputs

1.2 Use Case 2

Use Case 2 (UC2) aims at building an easy to use and portable data reduction pipeline of archived LOFAR observations for astronomers. This requires an easy to use web interface in combination with containerized workflows for portability. From the web interface, the astronomer can choose both the observations to be processed and the type of processing (pipeline) to be performed. Once the data and the processing are identified, the former are moved to the PROCESS infrastructure for processing and the final results (images) are sent back to the astronomer as URLs or shown as thumbnails in the web application.

The entire workflow is now implemented and can run on the PROCESS infrastructure. LOFAR data processing software has been containerised using Singularity. Initially, only the part performing direction-independent calibration was containerised. Later on, the second part performing direction-dependent calibration has been added. For the web interface, we reused the web portal developed for LOFAR in the EOSC pilot project[15]. The frontend has been modified to allow the selection of both the calibrator and the target observations. The backend for submitting computations has been extended to support submissions to the REST API of the IEE. The submission mechanism now supports both - xenon-flow[16] and IEE, and uses the common workflow language[17] (CWL) for workflow definitions. The data staging on the LOFAR long term archive (LTA), the transfer to the computing sites, and the results download are implemented using LOBCDER data services. Screen captures of the web portal are shown in Figure 7.

---

13 http://gradcam.cloudcv.org/

14 https://arxiv.org/abs/1710.11063

15 https://github.com/EOSC-LOFAR/ltacat

16 https://software.process-project.eu/software/xenon-flow

17 https://www.commonwl.org/

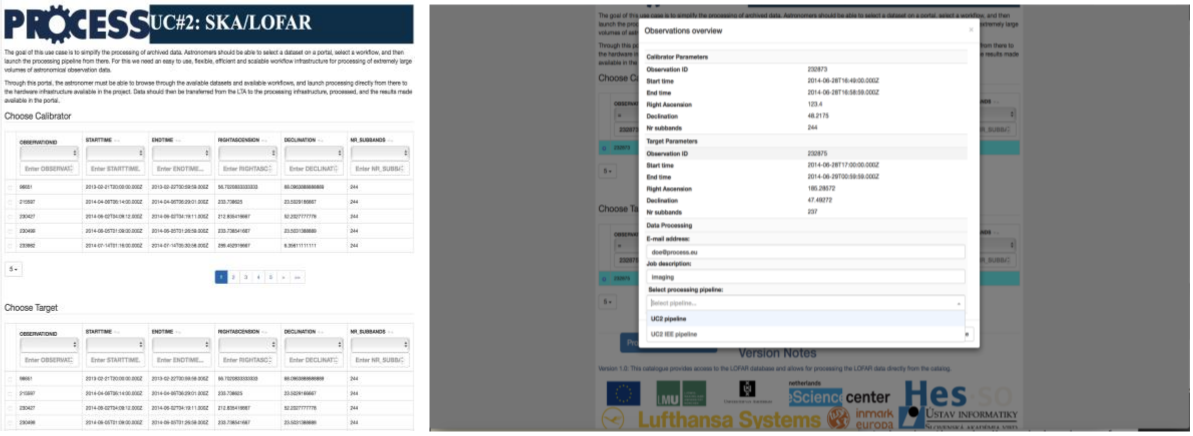

Figure 7: LOFAR web interface

For testing purposes, we use the same observational data as in D2.2 to show the evolution of the pipeline implementation. This data consists of a 25GB dataset (L232873) for the calibrator and a 433 GB dataset (L232875) for the target source. The UC2 pipeline implements a typical LOFAR data reduction pipeline with various pre-processing steps, a direction-independent calibration step and a direction-dependent calibration and imaging step which generates the analysis-ready images for the astronomers. Figure 8 shows images generated after each of these steps. We observe that there is a clear improvement in image quality as we go from uncalibrated data to direction-independent calibrated data to direction-dependent calibrated data, from left to right, upward then downward.

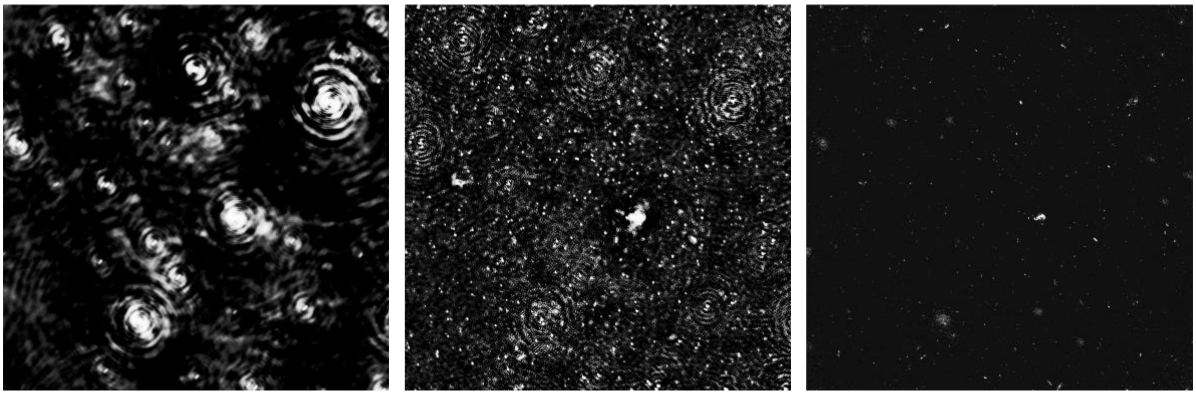

Figure 8: UC2 data reduction pipeline images

1.3 Use Case 3

Use Case 3 aimed at supporting innovation based on global disaster risk data in close collaboration with UNISDR that was recently renamed "UN Office for Disaster Risk Reduction (UNDRR)". As reported in deliverable D2.2, the PROCESS consortium regrets that the direct collaboration with UNISDR/UNDRR could not produce the expected outcome, since the UNISDR organisation had dissolved in its original form.

The goal of Use Case 3 is to enable more efficient data management was covered within the project by the enhanced data service around LOBCDER as reported in D4.5. Through those integrated and containerized services, it is possible to access and manage data according to the FAIR principles. This solution was also used in Use Case 5 to enable user downloads of the output data sets. Thereby, a containerized web portal allows access to data stored on any resource location integrated into the PROCESS ecosystem.

The former UNISDR data portal including two Terabytes of data sets was efficiently ported to a LOBCDER container solution on the PROCESS used infrastructures and is accessible online[18].

The PROCESS partners will continue to promote the usage of this efficient and lightweight solution to other communities as one additional building brick in combination with the data service. Data management and access is still one of the main challenges towards Exascale. Our solution, integrated in a comprehensive framework, is planned to be discussed with the EUDAT[19] project, to find synergies and potential use cases in the future.

1.4 Use Case 4

Use Case 4 aims at developing a software that generates revenue-optimal prices for the “first bag” ancillary in the airline industry and proves the algorithm’s superiority compared to today’s static price approach. The “first bag” ancillary is the additional purchase of a bag / suitcase which is checked-in and transported in the cargo compartment of the aircraft. As described in Deliverable D2.2, our remaining business challenge was the development of the revenue simulation (for proving the algorithm’s superiority) as well as the technical challenge of the integration of the use case in the PROCESS environment and the testing of the performance of the pricing calculator.

To finalize the business tasks, we developed a revenue simulator which creates a set of random passengers whose properties are distributed along the same distribution as used in the generation of the set of artificial passengers. The probability distribution is in this case a random distribution given by a probability tree. According to this tree, several parameters of the passenger are randomly drawn, e.g. traveller type (business or leisure), age and most importantly time of ticket purchase with respect to departure date (how many “days before departure” the ticket was purchased).

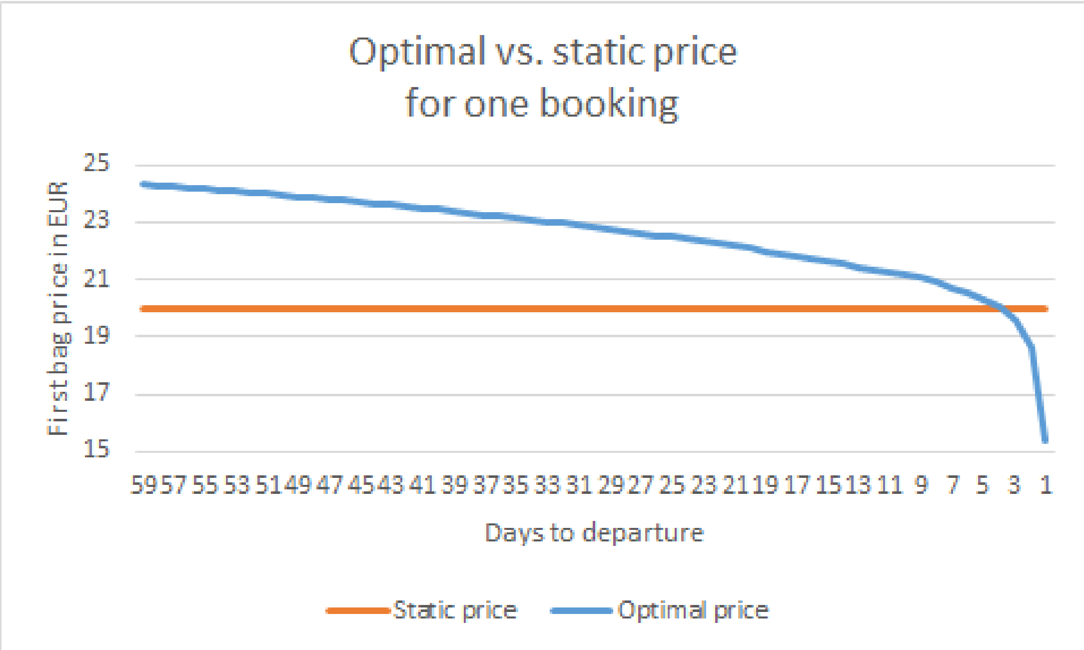

Figure 9: Depiction of the development of the optimal price for the first bag ancillary for one booking

Then the pricing calculator determines for each of the remaining days before departure an optimal price for the first bag ancillary which is offered to the passenger. Figure 9 shows an example of the development of the optimal price over time w.r.t. the days before departure of the flight.

---

[18] https://gar.mnm-team.org/

[19] https://eudat.eu

The figure shows that at the time of booking (59 days before departure) the optimal price is near 25 EUR. However once moving closer to departure, the optimal price declines, since we can obviously only sell this ancillary until the flight departs. However, for a practical application of the algorithm we have to address that the price cannot always become cheaper each day, hence this will quickly circulate among passengers and lead to everybody booking their first bag ancillary on the day of departure.

After the optimal price is calculated for every day before departure, we can calculate the expected revenue from this passenger for the optimal prices and for an offer with today's static price for the first bag, based on the known likelihood of a passenger to buy the first bag ancillary at a given day before departure.

Table 1 shows the results of the expected revenue of the optimal price versus the static price for a sample size of 1000 passengers, resulting in about 12,000 offers of the first bag ancillary:

Table 1: Expected revenue increase from static to optimal price for first bag ancillary in a simulated passenger set of 1000 passengers

Expected revenue increase from static to optimal price	**Occurrence for number simulated passengers (out of 1000)**	**Occurrence as percentage of simulated passengers (out of 100%)**
< 1%	15	2%
1 - 3%	862	86%
3 - 5%	82	8%
> 5%	41	4%

In all cases we found that the optimal pricing increases the expected revenue with the passengers. In 86% of the passengers, the revenue increase was between 1-3%. In Euro, this revenue increase accumulated to 430 EUR for all 1,000 passengers combined. This may seem like a small amount, but extrapolated to a big airline with 100 million passengers per year, accumulates to a 6-digit revenue increase with just one single ancillary.

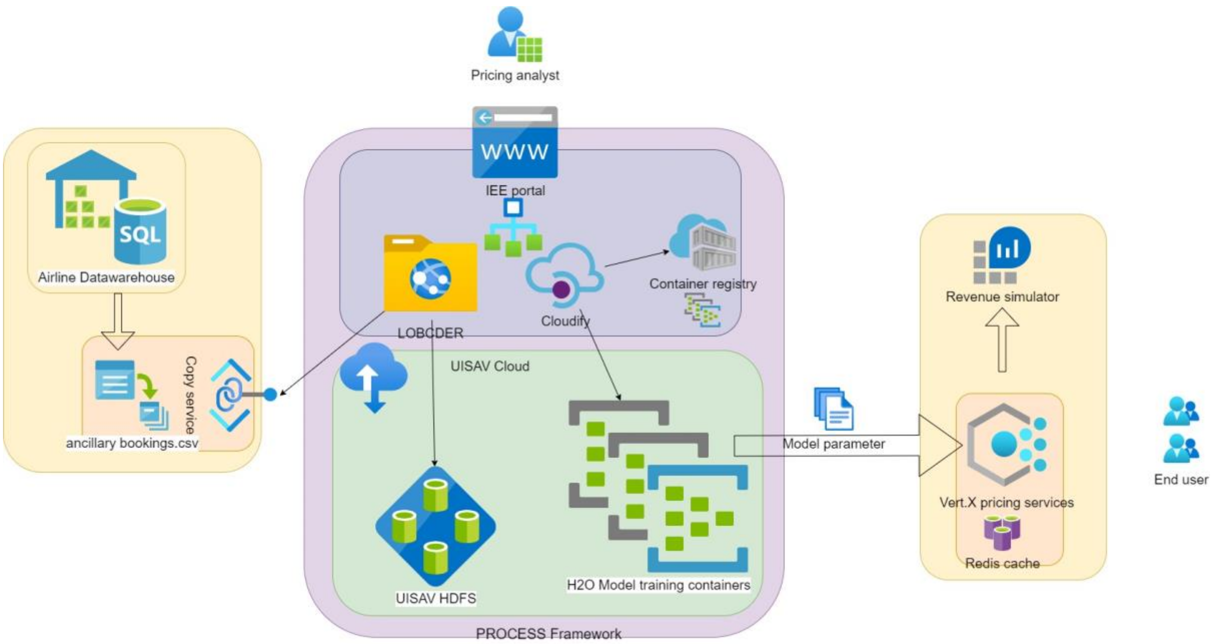

Figure 10: Use Case 4 integration

For the remaining technical challenges, we integrated the use case software into the PROCESS environment as shown in Figure 10. The data used for training the pricing model originates from the airline's data warehouse. Since there is no direct access to the airlines' data warehouses, some extraction steps are required before making the data available to the PROCESS framework. Due to this requirement, an intermediate step was introduced: the data extracted from the data warehouse is stored on a MS Azure storage. The LOBCDER data stage-in mechanism takes the data from these Azure storage locations and puts them into the environment, where the actual processing takes place.

The trained regression model is exported from the model training environment and is available for download through the WebDAV interface.

The last component in the above depicted high-level architecture is the pricing service. This service requires a trained model to serve as a basis for calculating and answering real-time pricing requests. Based on the request parameters a prediction is made with the regression model and a revenue optimal price is calculated with an optimization algorithm. Answering the request including prediction and calculation takes less than three seconds, which satisfies the use case's requirements. Through the asynchronous non-blocking architecture of the pricing service we were able to serve some hundred request per seconds on commodity hardware, which extrapolates to some 100 million request per day and therefore satisfying the requirements of the airlines. With further increase in CPU cores and memory of the hosting virtual machine, this throughput can even be scaled out further.

1.5 Use Case 5

Simulations and outcomes of earth science and observation applications increased their impact for the daily life dramatically within the last years. Monitoring the usage of resources for agricultural purposes impacts farmers directly and indirectly costumers and more severely and importantly our climate. Science and industry work hand-in-hand to enhance those simulations and the validation with actual observations.

This Use Case bases its simulation on regular updated Copernicus datasets, covering the entire globe. Together with real observations the target application and can simulate, for example, the growth of a specific seed for any time series on the entire earth by calculating the leaf area index (as displayed in Figure 11).

Necessary computations and data storage exceed more and more the available resources for those long running simulations. PROCESS enabled any end-user to deploy an instance of such an earth observations tool on a closed-off target system. Thereby, the growth and distribution of seeds could be simulated including the configuration of several major variables of the computation.

The closed source nature of this use case presented a unique requirement for the PROCESS ecosystem that could be realised by implementing a generic and configurable API-container, deployed at the use case site as an intermediary between the PROCESS infrastructure, namely the IEE, and the SME's closed source application.

This API enables the integration of any closed source application with PROCESS and enables job submission and status control of the PROMET HPC workflow, making job management easier for the use case owner and providing new scaling opportunities for the use case. Through the APIs integration with LOBCDER it is further possible to gather the outputs of PROMET's computation and make them available for end-users through the IEE.

The development of this container has been finalized and reported in D4.5. The validation did not show overhead since all requests are directly forwarded to the HPC nodes. It has been publicly released as part of the PROCESS software release [20].

[20] https://github.com/process-project/UC5-API

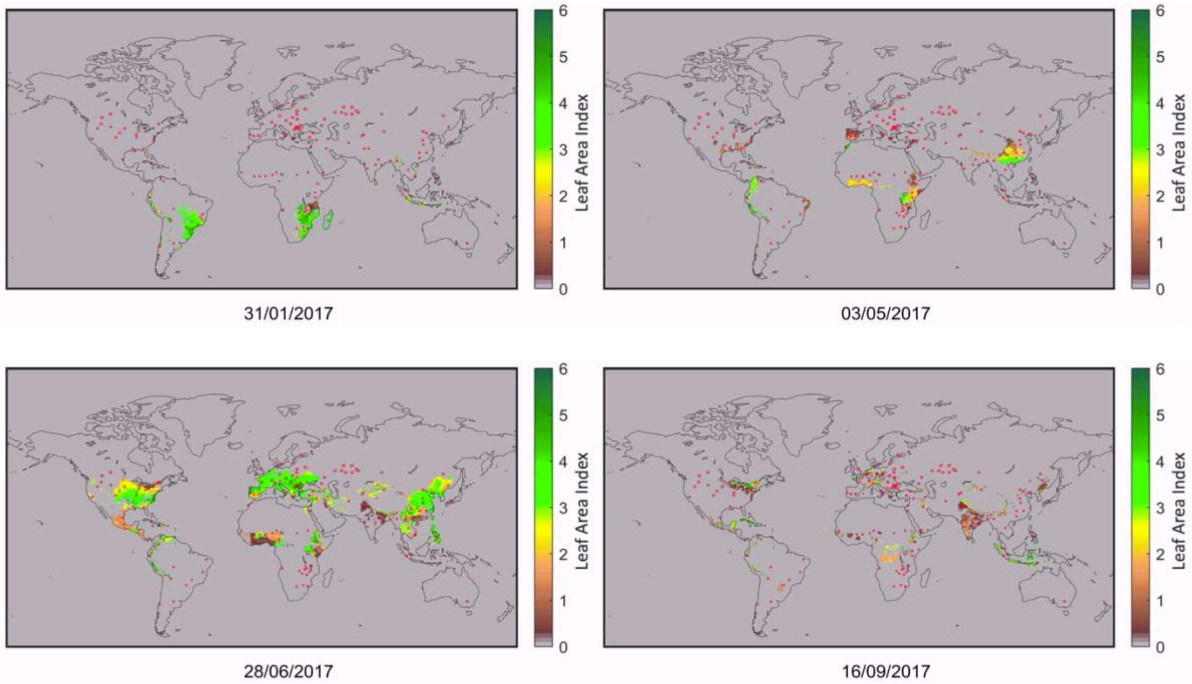

Figure 11: Use Case 5 Global high-resolution ensembles: Example Maize LAI

Several of those computations were successfully deployed by the IEE on the LRZ site. As part of a large calculation the development of maize was simulated for the year 2017. In Figure 11, the leaf area index (LAI) is shown for January, May, June and September 2017. The growth can easily be seen from south to north, where the green colours indicate a large LAI. The simulation includes images for each time step of 1 hour and is therefore a very complex and long running simulation.

Such simulations can now be easily deployed by anybody from within the PROCESS web interface. Through the abstraction and added layer upon, the end-user has no idea of the complex workflow started by him or her with only one click. This solution developed by PROCESS enables a broad usage of such closed source applications, without exposing the source code.

1.6 Summary

In this deliverable, we reported the final status of all PROCESS Use Case prototypes and applications. Each Use Case was able to leverage the PROCESS infrastructure and benefit from its scaling potential. The variety of the different presented Use Cases also highlights the modular building bricks of PROCESS, making it possible for different Use Cases to pick and choose the necessary components that could most benefit the users while keeping overall complexity low. Each Use Case prototype was publicly released within the PROCESS software release. Use case developers will continue their development using the PROCESS ecosystem beyond the end of the project. Each use case prototype release was also developed with modularity and reusability in mind and documented within, to enable new user communities and similar Use Cases to benefit from these releases by adapting them to their own needs.

2 Data Management Report

This section gives an overview, of the way PROCESS implemented the Data Management Plan (DMP) presented in D1.1. This deliverable functioned as a starting point regarding each data generation and publication throughout the project.

As included in the initial DMP, the PROCESS services and use cases were never meant to produce data sets. Both, services and pilot applications rather made use of data sets or developed algorithms and learning models. Therefore, the main part of the data management includes publications of software and algorithms.

2.1 Data Management in Services

The PROCESS services consist mainly of source code modules, which are already released and documented in the Research Software Directory[21]. Each developed software was released under an open source licence and can be reused by any other service ecosystem, as a stand-alone version or in conjunction with more services and a comprehensive instance of the PROCESS ecosystem (compare D8.4).

2.2 Data Management in Use Cases

All five Use Case prototypes showcasing the PROCESS service ecosystem demanded and had different challenges regarding their data management. In D1.1 we presented an *a priori* assessment of those key challenges. Throughout all submitted deliverables, focusing on the architecture and use cases, we reported the progress of the use case specific processes, how their final and intermediate results are generated. Following, the overall data usage and publication is summarised per Use Case.

Use Case 1

This Use Case used only already publicly available data sets (D1.1 Table 3). All developed machine learning algorithms to distinguish the optimal learning models were released under the MIT license and the source code[22] is available online.

This kind of publishing is in line with the planned publishing approaches described in D1.1.

Use Case 2

The LOFAR related Use Case did not produce any data sets, it made use of observations stored and published in the LOFAR Long Term Archive (LTA[23]). Access to these data sets was enhanced through several software modules and made it much easier for any astrophysical scientist to process those observations.

The main developed modules are the *LOFAR LTA One-Click-Processing Frontend*[24] and *The Brane Framework*[25], which both contributed to a generic and easy data access strategy.

21 https://software.process-project.eu

22 https://software.process-project.eu/software/camnet

23 https://lta.lofar.eu/

24 https://software.process-project.eu/software/lofar-lta-one-click-processing

25 https://onnovalkering.github.io/brane/

Use Case 3

Supporting the data management of UNISDR data sets could not be completed successfully, since the UNISDR organisation had dissolved in its original form, as reported in D2.2 Section 4. The only asset of this Use Case is the data set access portal (https://gar.mnm-team.org/) released as a data service module within the PROCESS ecosystem[26].

Use Case 4

This industry related Use Case bases their production system on confidential data sets, not accessible to PROCESS. To include the model generation into the PROCESS ecosystem, an artificial data generator was implemented and published[27], which produced artificial data sets as input for the model training. To reproduce these artificial data sets, the corresponding algorithms have been released according to the DMP.

Use Case 5

This closed source Use Case owned by a SME had unique requirements to protect the used software as an asset. It was initially planned to include the data management between the Copernicus data sets and the source application as a pre-processing module. Since this could not be fulfilled due to the closed source character of the Use Case, the only managed data includes the published API[28] and output data sets, produced by the applications. These data sets are accessible through an instance of the data set access portal developed for Use Case 3. The data set contains a time series of images of the processed simulation.

2.3 Summary

The initial and updated DMP support each partner in their decision, how and if at all their produced assets should be published. The PROCESS DMP did not need to be changed, since the comprehensive analysis at the beginning proved to be valuable. All data management actions were carried out according to the DMP.

[26] https://github.com/process-project/UC3-Portal

[27] https://github.com/process-project/UC4-AncillaryPricingDataGenerator

[28] https://software.process-project.eu/software/uc5api

3 Architecture Evaluation and Dissemination Report

The architecture of the PROCESS project evolved during the whole lifetime of the project. Each version of the architecture was improved by reviewing all requirements coming from different user communities (such as medicine, radio astronomy, airline revenue management, etc.). The main challenge was to propose an architecture that is suitable for all of them. This section therefore gives an overview of the final PROCESS architecture, summarising previous deliverables. Next, we present our approach for the evaluation of the architecture and present the related dissemination activities geared to reach new audiences that would benefit from this approach.

3.1 Final architecture

The requirements of the use cases are the foundation for the design of the PROCESS architecture. These requirements were analysed exemplary for the 5 PROCESS Use Cases and generalised. These requirements can be divided into three main groups:

1) virtualization requirements,
2) data requirements, and
3) computing requirements.

The main outputs of the architecture design are:

1) the reference exascale architecture, and
2) PROCESS architecture.

While the reference architecture characterizes key attributes and properties that have to be handled by every scientific application using exascale data and computations. The PROCESS architecture describes the realised version of architecture, deployed as the PROCESS platform.

- In D4.1, the design process is initialized. From the initial requirement analysis and the common conceptual model together with the overview of the technological state of the art, the initial architecture is developed.
- In D4.3 (after validation of the first PROCESS platform prototype), new requirements are identified which are coming from the experience and knowledge gained in implementing the PROCESS platform components. This analysis forms the foundations of the reference exascale architecture and its key features (such as distributed file system spanning a large number of heterogeneous computing infrastructures, the adoption of the virtualization and modular micro-service infrastructure).
- In D4.5, the final architecture is presented. The main improvement is the redesign of the computing infrastructure in order to optimize computing resources management. The final architecture takes into account the concept of data transfer nodes. Last but not least, the architecture describes the integration of the LOFAR portal as well as with the European Open Science Cloud.

3.1.1 Evaluation approach

The aim of the validation process is to have a mechanism to judge the PROCESS solution as a whole and ensure the use case requirements are met. This is achieved by checking the architecture and deploying and running the use cases on the PROCESS platform. A tool in achieving this goal consists of checking whether the platform reaches a given technology readiness level (TRL) by assessing its components and their interactions. As stated in D4.5, the PROCESS platform has reached TRL 5.

As stated in D4.3, the validation process follows a two-step approach. First, by assessing whether functional requirements of each use case are met by the PROCESS architecture. This is described in Section 4.1. Secondly, by measuring and analysing the operational overhead of the PROCESS platform which helped us to assess the scalability of our solution. In general, the overhead of the PROCESS platform stays within reasonable limits making it a potential exascale solution. It is important to note here, that PROCESS platform and the evaluations conducted during the PROCESS project were limited to the software stack. Further optimizations at the infrastructure (hardware level), beyond the scope of this project, can be performed to reach even higher performance. We have for instance pointed out the use of Data transfer nodes (DTNs). Further details about PROCESS platform and performance evaluations are given in Section 4.2. The details of the validation of the first prototype of the PROCESS platform are reported in D4.3 where we show that most of the use case requirements were satisfied using a combination of services composing the PROCESS platform. In this document in Section 4, we re-assess unmet UC requirements in the light of component/service evolution and re-evaluate the platform performance of the final prototype. The PROCESS platform reached TRL 6 for the final prototype.

3.1.2 Dissemination of the architecture

We organised two half-day workshops Platform-driven e-infrastructure innovations (EINFRA) at the IEEE eScience International Conference: 1) the eScience 2018 workshop in Amsterdam, The Netherlands and 2) the eScience 2019 workshop in San Diego, California, USA, where exascale projects and use cases were invited to join. Because of the COVID-19 pandemic restrictions, we could not plan a third workshop within the lifetime of the project, instead, we decided to organise a special issue in the journal Computing and Informatics (indexed in Current Contents).

Following is the list of the international events, in which the PROCESS work has been presented:

1. Bobák, Martin - Belloum, Adam S. Z. - Nowakowski, Piotr - Meizner, Jan - Bubak, Marian - Heikkurinen, Matti - Habala, Ondrej - Hluchý, Ladislav. Exascale computing and data architectures for brownfield applications. In 14th IEEE International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2018), pp. 461-468, ISBN 978-1-5386-8097-1. Huangshan, China, July 2018.
2. Martin Bobák, Ladislav Hluchy, Adam Belloum, Reginald Cushing, Jan Meizner, Piotr Nowakowski, Viet Tran, Ondrej Habala, Jason Maassen, Balázs Somosköi, Mara Graziani, Matti Heikkurinen, Maximilian Höb, Jan Schmidt. Reference Exascale Architecture. In Proceeding IEEE 15th International Conference on eScience (eScience) 2019 : Workshop on Platform-Driven e-Infrastructure Innovations (EINFRA), San Diego, California, USA, IEEE 2019 p. 479-487. IEEE Catalog Number: CFP1978F-ART. ISBN: 978-1-7281-2451-3. DOI: 10.1109/eScience.2019.00063.
3. Reginald Cushing, Onno Valkering, Adam Belloum, Cees de Laat. Towards a New Paradigm for Programming Scientific Workflows. In Proceeding IEEE 15th International Conference on eScience (eScience) 2019 : Workshop on Platform-Driven e-Infrastructure Innovations (EINFRA), San Diego, California, USA, IEEE 2019 p. 604-608. IEEE Catalog Number: CFP1978F-ART. ISBN: 978-1-7281-2451-3. DOI: 10.1109/eScience.2019.00083
4. Valkering O, Belloum A. Privacy-Preserving Record Linkage with Spark. In2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID) 2019 May 1 (pp. 440-448).
5. Höb, M., PROCESS -- PROviding Computing solutions for ExaScale challengeS (Poster), In ICCS 2019 Proceedings and Book of Abstracts, Faro, Portugal, June, 2019.
6. HLUCHÝ, Ladislav - BOBÁK, Martin - MÜLLER, Henning - GRAZIANI, Mara - MAASSEN, Jason - SPREEUW, Hanno - HEIKKURINEN, Matti – PANCAKE-STEEG, Jörg – SPAHR, Stefan – FELDE, Nils – HÖB, Maximilian – SCHMIDT, Jan - BELLOUM, Adam S.Z. - CUSHING, Reginald - NOWAKOWSKI, Piotr – MEIZNER, Jan - RYCERZ, Katarzyna - WILK, Bartosz – BUBAK, Marian - HABALA, Ondrej – ŠELENG, Martin - DLUGOLINSKÝ, Štefan – TRAN, Viet – NGUYEN, Giang. Heterogeneous Exascale Computing1. Chapter 5. In Recent Advances in Intelligent Engineering: Springer, 2020, p. 81-110. ISBN 978-3-030-14349-7.
7. BUBAK, Marian - MEIZNER, Jan - NOWAKOWSKI, Piotr - BOBÁK, Martin - HABALA, Ondrej - HLUCHÝ, Ladislav - TRAN, Viet BELLOUM, Adam S.Z. - CUSHING, Reginald - HÖB, Maximilian – KRANZLMÜLLER, Dieter - SCHMIDT, Jan. A hybrid HPC and Cloud platform for multidisciplinary scientific application1. In 6th International Conference on Supercomputing Frontier Europe 2020 - Warsaw, Poland.

3.2 Conclusion

Exascale computing and data are still facing theoretical and practical challenges. In PROCESS, we proposed a reference architecture which describes general concepts applicable to a wide range of scientific applications beyond the five use cases considered in PROCESS. To validate our proposal, we engaged with existing exascale communities in the context of the two workshops we organized during the lifetime of the project which helped us to collect feedback and comments on PROCESS results and methods. Finally, we organised a special issue in a scientific journal focused on exascale computing and data to disseminate and share the experience and the lesson learned during the project with the rest of scientific community. Last but not least, we proposed and implement the PROCESS architecture, described in D4.1, this architecture was updated twice according to the requirements coming from user's communities and service providers.

4 Validation Report

This section concludes the final validation of the PROCESS ecosystem. First, the fulfilment of the functional requirements of the five Use Cases is presented. Second, the validation of the PROCESS platform performance presents our approach to validate all components and verify their potential to scale towards exascale.

4.1 Functional use case requirements

In this section we revisit the functional requirements of the use cases, as originally defined in D4.1, updated in D4.3, and finalized in D2.3. We show that all requirements have either been met in the final PROCESS release, or dropped in one of the updates.

As of D8.2, most PROCESS components and services were already integrated and running as part of the production prototype, and the requirements for use cases 3, 4 and 5 had already been fully met. Use cases 1 and 2 had a small number of requirements still open, which we will discuss in more detail below.

4.1.1 Use Case 1

In Table 2 we provide an overview of the final requirements of use case 1, as reported in D2.2 (Table 1, pages 5 and 6). As of D8.2, there was a single unmet requirement for this use case: "*Support of Hovorod tool for distributed parallelization*". This requirement was met in D1.3, by porting the use case Layer 2 software to the cloud via Docker containers. Configurations of virtual machines with different GPU types were created and tested through the deployment of Docker containers with the required libraries, Horovod and OpenMPI. The obtained results are reported in Section 1.1 of this deliverable.

Table 2: Final requirements for use case 1

Initial (D4.1)	Updated (D4.3)	Final (D2.3)	Fulfilled
Support of Docker containers	Fulfilled and extended with Singularity containers	Support for both Docker and Singularity on Computing Sites	✓
Environments Manager that could guarantee a flexible building, deployment and management of multiple running applications	Guaranteed by the online IEE	Workflow Management for configuration, deployment and management of multiple application / use case executions	✓
Data Storage system that takes into account the high level of variation in the image formats across the different datasets and ensures flexibility and adaptation to a variety of datatypes.	Obtained thanks to an efficient data-pre-processing pipeline, the distributed data system and the integration with LOBCDER	Integration with LOBCDER to access all the data Storage Centres	✓
		Support for multiple pipelines / pre-processing pipelines within the workflow management	✓
User-friendly environment for uploading, locating and transferring of the data and download of outputs	Data transfer by SCP protocol.	Support of SCP and GridFTP.	✓

		Efficient and user-friendly data upload, transfer and download obtained by integration of LOBCDER and WebDAV	✓
		User-friendly interface for data access through the workflow management	✓
Support of the set of common tools for machine learning and deep learning	Supported on the HPC centres TensorFlow, Keras, Theano, PyTorch, Openslide	Support for both Docker and Singularity on Computing Sites	✓
		Containerize use case	✓
Access to GPUs for network training	Parallel distributed dense linear algebra and multi-GPU settings	Provide Access to GPU Resources of Computing Sites	✓
		Support GPU Usage in Containers	✓
Security and privacy: anonymize data before use	Fulfilled by the use of publicly available data	Compliance with EU GDPR	✓
Support of Horovod tool for distributed parallelization	Horovod available on local computation centres	Integration of Horovod within the Cloud	✓

4.1.2 Use Case 2

In Table 3,we provide an overview of the final requirements of use case 2, as reported in D2.3 (Table 2, page 6). During the writing of D8.2, two requirements of use case 2 were still partly open.

First, as part of the first two UI requirements shown in Table 3, the use case specific UI based on the EOSC pilot project needed to be properly integrated with the IEE. This work has been completed in this deliverable, allowing the UC2 interface to communicate seamlessly with the IEE through a REST-API.

Second, for maximum flexibility and throughput for the analysis pipelines, UC2 needed to support multisite execution. This has been achieved by extending the UC2 portal to allow the user to select different destination sites when stating an execution of a pipeline. Within the PROCESS testbed, however, the Prometheus supercomputer located at CYFRONET is the only available infrastructure capable of running the UC2 containers. We could therefore not fully evaluate this feature.

During the project, several initial requirements were dropped. First, the “CVMFS client to distribute the containers” and “PiCaS installation to distribute work” requirements were replaced, as these components were part of a previous effort to automate the running of UC2 analysis pipelines. It soon became clear that these components would be replaced by the IEE within the PROCESS framework.

The “Support DTNs to speed up transfers” requirement was added in D2.3 to improve the data transfer performance between the LTA archive and the PROCESS compute infrastructure. Despite our efforts, however, complexity of setting up such data transfer networks proved to be beyond this project's capabilities. As it was not part of our original scope, we therefore reverted to the original solution.

Table 3: Final requirements for use case 2

Initial (D4.1)	Updated (D4.3)	Final (D2.3)	Fulfilled
A user-friendly environment for selecting the data and the workflows	Reuse of a web interface from EOSC pilot project and AA-Alert projects	Community portal integration for Workflow management	✓
Mechanisms for launching workflows from front-end	Mechanisms integrated into the web interface	Community portal integration for Workflow management	✓
Support for Docker or Singularity containers	Fulfilled by site admins installing Singularity or Docker	Support for both Docker and Singularity on Computing Sites	✓
		Containerize use case	✓
GridFTP for downloading the measurement sets (MS)	Provided as data service by LOBCDER	Support of SCP and GridFTP.	✓
VOMS client to manage GridFTP access	Provided through tool installation and LOFAR VO affiliation	VOMS client to manage GridFTP access	✓
CVMFS client to distribute the containers	Replaced with use of Singularity hub	Replaced	Not applicable
PiCaS installation to distribute work	Replaced with a work orchestration framework such as Kubernetes	Replaced	Not applicable
A sufficiently fast network for downloading the MS	Standard networks with GridFTP; DTNs and alternatives still work in progress	Support of SCP and GridFTP	✓
		Support DTNs to speed up transfers	dropped
Sufficient storage space for the MS, temporary data and output images	Fulfilled with extra scratch space on tested computing sites	Sufficient Storage Resources on Computing Sites	✓
		Provide Access to Storage Resources on Computing Sites	✓
CPU and later GPU clusters for data pre-processing with fast access to data	Use of standard compute clusters; use of GPU is work in progress	Provide Access to GPU Resources of Computing Sites	✓
		Provide Access to CPU Resources of Computing Sites	✓

4.1.3 Use Case 3

In Table 4, we provide an overview of the final requirements of use case 3, as reported in D2.3 (Table 3, page 8). All requirements for this use case have been met since the first prototype described in D4.3. While some requirements were added or adapted during the project, none were dropped.

Table 4: Final requirements for use case 3

Initial (D4.1)	Updated (D4.3)	Final (D2.3)	Fulfilled
	Easy-to-setup and use and extensible UI with data discovery functionality	Containerized Portal with easy configuration available	✓
	Support for multiple indexing/curation approaches	Interfaces to connect to different local and mounted file systems and remote repositories via SCP	✓
		Support for both Docker and Singularity on Computing Sites	✓
		Containerize use case	✓
	Integrated Data Service connected to the workflow management system	Modular Data service	✓

4.1.4 Use Case 4

Table 5 provides an overview of the final requirements of use case 4, as reported in D2.3 (Table 4, page 8). All requirements for this use case have been met since the production prototype described in D8.2. While some requirements were added or adapted during the project, none were dropped.

Table 5: Final requirements for use case 4

Initial (D4.1)	Updated (D4.3)	Final (D2.3)	Fulfilled
The system needs to be capable to deal with the large passenger data sets that airlines generate.	Data store for < 100TB Provided by Prometheus cluster	Sufficient Storage Resources on Computing Sites	✓
		Provide Access to Storage Resources on Computing Sites	✓
		Provide 100TB Storage Resources	✓
Scalable architecture which has the potential to: a. handle large amount of data b. handle data from different sources c. handle high volume of requests per day	Use of cloud infrastructure managed by Cloudify integrated into IEE.	Provide Access to Cloud Resources	✓

d. provide quick response times e. be extensible in terms of continuously increasing data as well as increase in parallel requests being sent.			
		Integrate Cloud Resources with Workflow Management	✓
		Provide Sufficient Cloud Resources	✓
Establishing a consolidated data structure on which further statistical processing can be performed		Provide Access to Storage Resources on Computing Sites	✓
Processing of ongoing data streams to keep the consolidated data structure up-to-date (i.e. learning new data behaviour into reference data).	Use of cloud infrastructure managed by Cloudify integrated into IEE.	Provide Access to Cloud Resources	✓
		Integrate Cloud Resources with Workflow Management	✓
	Support of streaming data in addition to static data; use of H2O and Sparkling Water and Docker containers Provided through containerization.	Support for both Docker and Singularity on Computing Sites	✓
Distributed computing fundamentals based on the Hadoop ecosystem.	Use of cloud infrastructure managed by Cloudify integrated into IEE.	Provide Support for HDFS	✓
Provide data storage for at least two years' worth of historical data. Provide data storage for derived model parameters. Process two years' worth of historical data with machine learning algorithms	Data store for < 100TB Provided by Prometheus.	Sufficient Storage Resources on Computing Sites	✓
		Access to Storage Resources on Computing Sites	✓
Provide scalability to respond to hundreds of million pricing requests per day.	Use of cloud infrastructure Provided by Cloudify integrated into IEE.	Provide Access to Cloud Resources	✓
Software requirements Hadoop, HBase, Spark, Tensorflow etc. For the business side, common machine learning libraries have to be available. We need access to the Hadoop file system as well as to the HBase datastore.	Support for Hadoop/HDFS, HBase, Spark, TensorFlow provided mostly through containerization; HDFS is not containerized, but it is accessible through adaptor which is part of a micro-infrastructure.	Provide Support for HDFS	✓

		Support for both Docker and Singularity on Computing Sites	✓
		Containerize use case	✓
The software is supposed to be run at airlines in the European Union. Therefore, it has to comply with the "EU General Data Protection Regulation", if applicable.	Compliance with EU GDPR In the first stage we rely only on generated data, no real personal data is used.	Compliance with EU GDPR	✓
The software needs to be deployable on sight at the customer's cloud service, for example Microsoft Azure.	Use of cloud infrastructure Provided by Cloudify integrated into IEE.	Provide Access to Cloud Resources	✓
		Integrate Cloud Resources with Workflow Management	✓
		Support for both Docker and Singularity on Computing Sites	✓

4.1.5 Use Case 5

Table 6 provides an overview of the final requirements of use case 5, as reported in D2.3 (Table 5, page 10). All requirements for this use case have been met since D4.5.

The PROCESS-Copernicus-Adapter requirement has been discarded, since the pre-processing and data set import is done within the closed source application. Also, the Python development environment was discarded, since the Use Case did not provide any direct access to the source code.

Table 6: Final requirements for use case 5

Initial (D4.1)	Updated (D4.3)	Final (D2.3)	Fulfilled
PROCESS-Copernicus-Adapter	No Adapter needed	Discarded	Not applicable
PROCESS-PROMET-PrePROCESSor	Generalized Proxy API for closed-source use cases	Generalized Proxy API for closed-source use cases	✓
PROCESS-PROMET-Adapter	Integration of LOBCDER with Proxy API for data stage-out	Integration of LOBCDER with Proxy API for data stage-out	✓
Access to HPC resources (Execution and Testbed / Deployment)	Access to HPC resources	Support use cases with restricted access to resources within Workflow Management	✓
Access to data storage	Access to data storage	Sufficient Storage Resources on Computing Sites	✓
		Provide Access to Storage Resources on Computing Sites	✓

Transfer Copernicus data sets to data storage	Data Service for data staging	Integration of LOBCDER with Proxy API for data stage-out	✓
Python development environment	Discarded	Discarded	Not applicable
	Support configuration of the execution parameters	Workflow Management for configuration, deployment and management of multiple application / use case executions	✓
		Integrate Workflow Management with Proxy API	✓
	Submit and activate workflow execution via API	Generalized Proxy API for closed-source use cases	✓
	Store PROMET output files for visualisation and download	Sufficient Storage Resources on Computing Sites	✓

4.2 Platform performance

This section provides the final validation of the performance of the PROCESS platform. The goal of this validation is to ensure that the PROCESS platform does not introduce an unnecessary overhead which restricts use case performance or scalability.

The work related to this validation has been performed over the course of the PROCESS project and has been reported on in detail in D3.3 and D8.2. The end result consists of a performance model of the overhead incurred by the PROCESS platform, and several aspects of the PROCESS testbed. We will provide a summary of the most important results here.

4.2.1 Performance indicators

Our performance measuring and modelling efforts are centred around the following definition of *runtime,* as defined in D3.3 (page 12):

Runtime = Overhead + Data Transfer + Scheduling + Execution Time

This *runtime* will be experienced by the user as the completion time of their workflow; how long do they need to wait to run it start-to-finish?

This *runtime* is determined by a number of factors: the time it takes to move data to and from the compute site (*data transfer*), the time the workflow needs to wait before the compute resources are made available (*scheduling*), the processing time itself (*execution time*), and the time it takes the PROCESS platform to orchestrate all of this (*overhead*). Each of these time components contributing to the *runtime* consist of one or more measurands which can be determined experimentally:

Overhead = T1 + T2 + T4 + T7
Data Transfer = T3 + T8
Scheduling = T5
Execution Time = T6

These measurands are defined in Table 2 of D3.3 (page 10). For convenience we replicate this table in Table 7 below:

Table 7: Summary of the various identified times, we will use as performance measurands

Tx	Name	Description
T1	Configuration	Time to configure the workflow for the application
T2	Deployment Strategy	Time to select appropriate storage and computing site
T3	Stage-In	Time to transfer data from source to selected storage site
T4	Container selection	Time to select specified container for the workflow from repository
T5	Schedule	Time the submitted job spends in queue
T6	Execution time	Time spent executing the job on the compute resource
T7	Stage-Out Strategy	Time to select appropriate storage site for output
T8	Stage-Out	Time to transfer result to storage site

Several of the defined measurands are not directly related to the PROCESS platform itself, but determined by external factors instead. Specifically, T6 (execution time), which is directly determined by the use case codes, T3 (stage-in) and T8 (stage-out), which are determined by the combination of the use case input and output data sizes, and the bandwidth available on the long distance network infrastructure of the PROCESS testbed, and finally T5 (Schedule), which is determined by the availability of compute resources within the PROCESS testbed and the scalability of the scheduling systems used (such as SLURM).

The overhead time component contains all other delays (T1, T2, T4, T7) which can be directly attributed to the use of the PROCESS services. For a detailed discussion on identification of the measurands we selected, we refer to D3.3, Section 2.

During the course of the PROCESS project, we have performed many measurements to get insight into the values of these measurands under different circumstances, using both micro benchmarks and complete use case workflows. These measurements were then used to create a performance model of the PROCESS platform and testbed overhead, which we will summarize below. We will not discuss the performance measurement for the individual use cases. These are reported in detail in D3.3 Section 4.2, and Section 1 of this deliverable.

4.2.2 Performance model

Using the various measurements performed during the PROCESS project, we have created three performance models to estimate the *overhead, data transfer,* and *scheduling* components of the *runtime*.

The *overhead* component is determined by the PROCESS platform itself, that is, the collection of services we delivered in this project. Therefore, we need to show that this platform indeed scales to the exascale problem sizes which are expected in the near future.

The *data transfer* and *scheduling* components are mainly determined by external factors which influence the performance of the PROCESS testbed, such as network bandwidth between storage and compute sites, the number of compute resources at these sites, and the scalability of the scheduling system used. Therefore, their performance is largely outside of our control. We can, however, use the performance models to estimate if these time components will become a bottleneck when scaling to exascale problem sizes.

As of June 2020, the #1 supercomputer in the Top500, the Japanese Fugaku[29] runs at approximately 0.4 exaflops using almost 159K nodes. We will therefore use our performance models to validate the

[29] https://www.r-ccs.riken.jp/en/fugaku/project/outline

performance of the PROCESS testbed and platform, up to 100K containers and 1 exabyte of data. For further details on the models we refer to Section 5 in D3.3 (pages 28-30).

The results for the *data transfer* and *schedule* and overhead models are shown in Figure 12, Figure 13 and Figure 14 below. These are also reported in D8.2 as Figure 1, 2 and 3, and are replicated here for convenience.

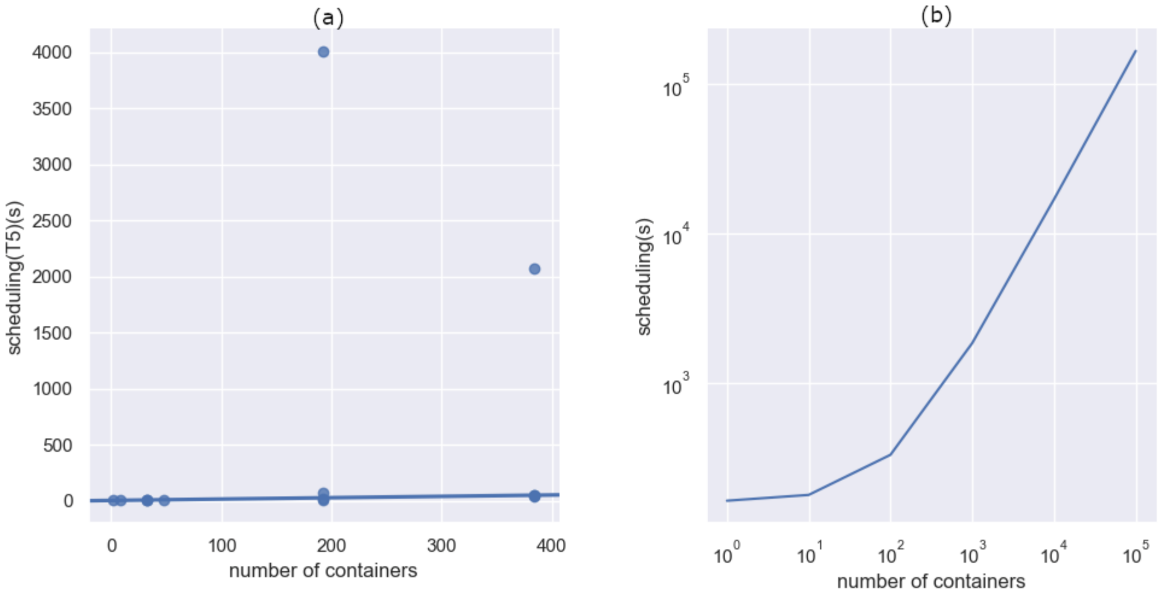

Figure 12: Scheduling overhead estimation at the exascale level

Figure 12(a) and Figure 12(b) show scheduling time estimations produced by our model. This is an overhead caused by the local scheduling infrastructure (such as SLURM) on the compute sites of the PROCESS testbed.

Figure 12(a) shows a relatively modest increase in scheduling time when the number of containers to be scheduled is increased to several hundred. This is expected, as these container counts are typical for current day systems. Therefore, scheduling systems have been optimized to handle such loads.

The projections obtained from the model for 100K containers is shown in Figure 12(b) at log scale. In contrast to the scheduling overhead for small container counts, the scheduling time is estimated to become large at very high container counts (in the same order as the container count itself). This is a clear indication of scalability issues in the scheduling system. Therefore, we expect that exascale data processing on such large number of nodes will present some challenges to current scheduling techniques and to require new approaches to scheduling.

Figure 13(a) and Figure 13(b) show data transfer time estimations produced by our model. In deliverable D3.3 (page 36), we showed that this *data transfer* is independent of the number of containers, but instead (and as expected) directly depends on the data size involved and performance of the network infrastructure available between test locations.

The linear model of the relationship is illustrated in Figure 13(a), shown below, for a small stage-in in the context of IEE. According to this linear model, it would on average take about 6 minutes (359 second) to transfer 10GB of data, at an average speed of about 28MB/s. While this transfer rate is sufficient for smaller scale data transfers up to a 100GB or so (which will take about an hour), it is already insufficient to stage in a full LOFAR observation of 16TB, which will take almost 7 days[30], thereby doubling the runtime of the workflow.

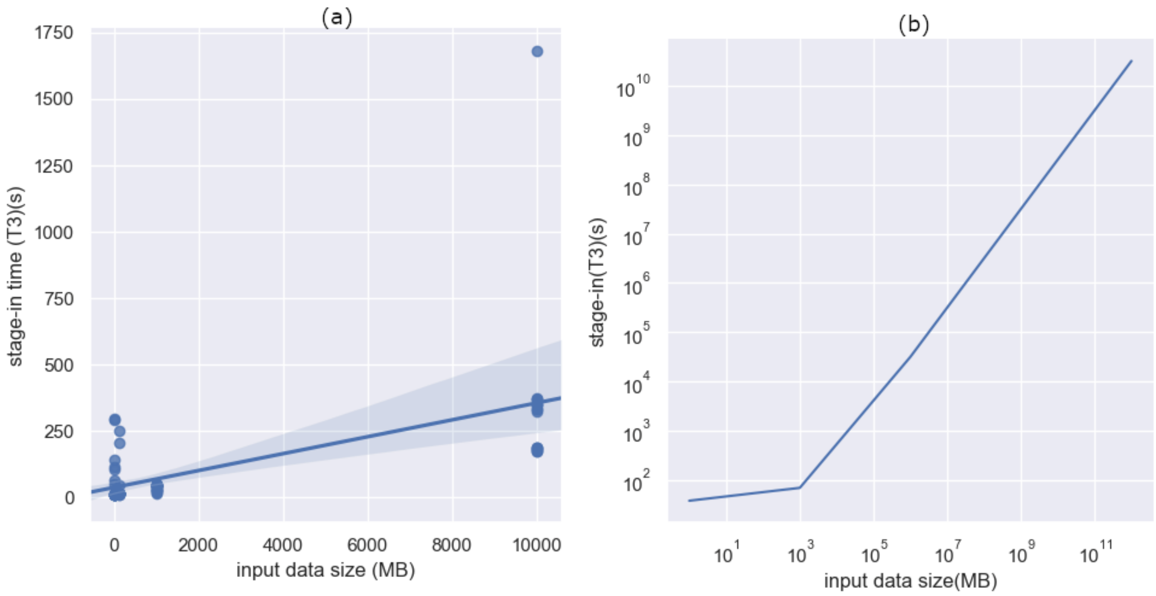

Figure 13: Data transfer estimation at the exabyte level

Using our model, we extrapolated hypothetical transfer times for sizes up to an exabyte, shown in Figure 13(b). As expected, because of the relatively low transfer rates in our testing environment, these data transfer time projections grow to extreme numbers for large datasets. For example, the transfer of 1 PB of data currently would take about a year. while transferring a full exabyte would take several centuries.

As discussed in D8.2 (Section 3.3), these data transfer issues can be solved by using a dedicated Data Transfer Network (DTN) infrastructure. Experiments by Geant in 2018[31] have shown that 100Gbit/s DTN networks are already feasible on a European and even global scale. At such transfer rates it would take about 21 minutes to transfer a single 16TB LOFAR observation and about 22 hours to transfer a petabyte. Transferring a full exabyte is still out of reach however, as it would require over 2.5 years. For this, a further increase of DTN network bandwidth into the TBit/s range will be required.

Finally, in Figure 14(a), we estimate the platform overhead up to several hundred containers. As the figure shows, the overall delay introduced by the various components is in the order of 22 seconds, which is well within acceptable limits when considering that data transfers and executions times typically takes hours or even days (for UC2)

In Figure 14(b), we use this model to extrapolate what the overhead would be as we approach an exascale size machine using an estimated 100K containers. This figure is using log scale due to the high container count. As the figure shows, with 100K containers the overhead is estimated to be about 1200 seconds or 20 minutes, which is still small when we put it in perspective with the large container count and often long execution times per container.

---

[30] Note that in D3.3 the transfer time of a LOFAR observation was erroneously estimated to be 18 years.

[31] https://www.slideshare.net/JISC/data-transfer-experiments-over-100g-paths

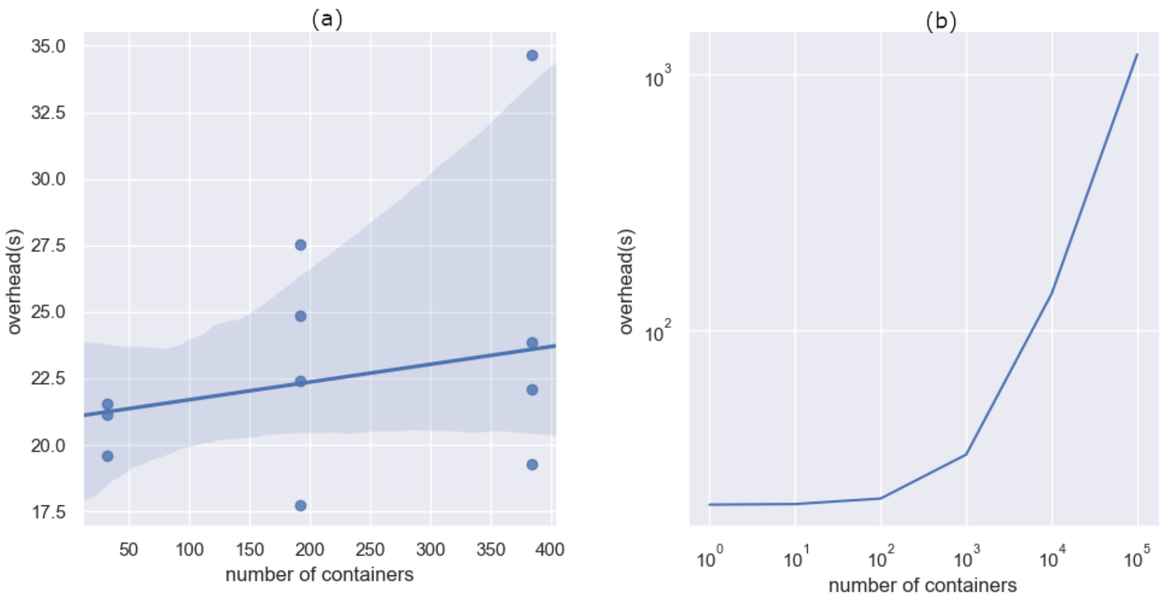

Figure 14: Platform overhead estimation at exascale machine size

4.3 Conclusion

In this section we have validated the PROCESS platform in two ways. First we have shown that the provided solutions meet the functional requirements of the use case as specified in D2.3. For all use cases a brief overview was given of the evolution of the requirements during the project lifetime, which requirements are met, and why certain requirements were dropped.

Next, we evaluated the performance and the scalability of both the PROCESS platform and testbed using the performance models developed in D3.3.

The estimates provided by these performance models indicate that scalability of the PROCESS platform is sufficient. While the overhead is expected to increase when large numbers of containers are used, it is not expected to become a major bottleneck soon.

The results are different for our data transfer and scheduling estimations of the PROCESS testbed. The local scheduling system (e.g. SLURM) is expected to become a bottleneck when approaching exascale size, and further scalability optimizations will be required for scheduling systems such as SLURM. However, as such scheduling systems are a major component of supercomputers, we expect these problems to be solved when the first exascale machines arrive.

The data transfer overhead is already significant in the current PROCESS testbed. The performance is insufficient to transfer the data sets required by UC2 within reasonable time. While solutions such as 100Gbit/s DTN networks exist and have been shown by Geant to be feasible on a European scale, these could unfortunately not be made available within this project. When extrapolating to exabyte scales, another tenfold increase in network bandwidth (to the TBit/sec range) will at least be required for acceptable data transfer times.

Both the local scheduling and data transfer performance are not directly within the scope of the PROCESS project. However, as external factors they do influence the performance of our use cases. Both issues are expected to be solved in future projects and can be combined with the PROCESS platform we have developed in a straightforward manner.

5 Innovation Potential Report

This section highlights the innovative aspects of each use case and further presents their potential to go beyond their current state of the art with the help of the PROCESS ecosystem.

5.1 Use Case 1

Deep learning frameworks and their large-data requirements have finally reached a point that requires collaboration with major data centres. Training time complexity and computing requirements have become a strong limitation in the development of successful applications of deep learning to the medical imaging field. Researchers are frequently forced to find alternative solutions to avoid the complexity of their models to reach the boundaries of the computing power they have available. This may affect the performance of these algorithms. Similarly, the size of the datasets makes data exchange prohibitive and development on multiple infrastructures is often discouraged in favour of an 'in situ' approach. To ease the development process and thus develop more powerful solutions for medical imaging the limits of computing power and data storage need to be pushed far beyond current boundaries.

PROCESS contributes to bring an infrastructure to researchers that makes it possible to develop novel pattern recognition and machine learning algorithms for the medical domain with improved flexibility, speed and more powerful systems. PROCESS allows researchers in the medical imaging community who are not experts in the field of HPC to access exascale infrastructures. This promotes the development of exascale deep learning models for analysis of large histopathology datasets. Highly complex models can be trained within the HPC infrastructure on massive datasets and made available for the digital histopathology community. Many architectures can be tested to extend learning to weakly annotated data. Intermediate results of computations that might be several orders of magnitude larger than the original data, are efficiently stored with PROCESS. Moreover, PROCESS provides an Evaluation-as-a-Service approach of the algorithms developed, which will be run in a sandboxed environment on the data that is not allowed to be shared.

As a result, we expect a decrease of the turn-around time of the experiments and of the development of the models for digital image analysis, thus allowing the development of applications otherwise computationally infeasible.

Access to HPC leads to better detection and classification of cancer stages, thus improving the treatment of medical conditions and consequently improving the quality of life of patients. The research community benefits from insights on the project via Open Access scientific publications.

The Evaluation-as-a-Service paradigm in the PROCESS project will provide additional interest for researchers by giving them the opportunity to run and evaluate their algorithms on the PROCESS HPC infrastructure on large data sets and subsets thereof in order to train their machine learning models, which can potentially improve the learning curve for scientific communities examining new techniques and services. At the same time, this approach also avoids the need for large data downloads and eliminates possible privacy issues that often come with medical data, as the data itself does not leave the hosting site's boundaries at any given time and can thus not be matched with other data, which can lead to security breaches. Beyond this, shipping the algorithm to the data where it is run on a specialized HPC architecture, can absorb the needs of research communities that require access to data to validate tools.

5.2 Use Case 2

Before PROCESS solutions, processing the LTA observations from visibility data to image pixels was notoriously difficult as it requires significant domain and programming skills.

With the PROCESS software ecosystem, a user-friendly Web portal which allows the selection of both the datasets and of the processing workflows is available to the astronomers. Following the launch of the workflows, PROCESS data services will transparently stage-in the observation data on the LTA and move them from there to the selected computing site where PROCESS compute services will start the analysis pipeline on them. Once available, the output images will be made available through the portal.

This simple and easy access to exascale infrastructure will further open up the LTA to the astronomy community, increase the scientific output of LOFAR, and serve as a stepping stone for SKA, which will produce even more data. Although we could not decrease the processing workflows execution time for lack of time, the ease of use and the straightforward nature of the Web portal relieving the user from the underlying complexity more than compensates for the lack of this feature.

5.3 Use Case 3

The initial roadmap for this use case included the development of an easy to deploy solution for data exploitation. This approach should base on a web front- and backend developed for UNISDR risk data sets. As already reported, within the project's lifetime the collaboration with the UNISDR department was not continued. Although the initial planned data sets were not available anymore, we continued to enhance the idea of the use case.

As reported in section 1.3, the PROCESS ecosystem with its capable data service was able to fully integrate the data exposure, listing and publication within a LOBCDER module. This easy configurable and deployable container was added as an additional building brick to the PROCESS ecosystem. Therfore, the use case already enabled a new service due its innovation potential. Since all requirements are met, and the use case potential is already adapted in a generalised module, the use case was of great benefit to the project and to similar future use cases, requiring the same functionality.

5.4 Use Case 4

From a technical point of view, the PROCESS software platform has proven to be capable to deal with the amounts of processing and data required by airline customers, making it a valid option for a real-life ancillary pricing solution.

The business side however has shown significant potential, since we were able to demonstrate a revenue gain of in average between 1-3% only for the first bag ancillary. This seemingly small revenue percentage gain however means an absolute revenue gain of several 100,000 EUR for just one single ancillary. And that is a very significant potential in a business with very low margins.

5.5 Use Case 5

The unique requirement of UC5 being a closed source application by SME, provided new opportunities for PROCESS by broadening the scope of support for potential new use cases. The developed API container for UC5 was generalized and released to the public, making it possible for new applications with similar requirements to be integrated with and to benefit from the PROCESS infrastructure and software stack.

Besides the new potential for use cases, this solution has also shown great potential to benefit the future of PROMET's modelling of agricultural observation, by making it possible be more flexible in ways to control and validate the behaviour of the modelling system and making submissions of multiple simulations easier to manage.

The new PROCESS tools make it possible to perform these multi-model simulations more efficiently by providing the ability to link observational data (and other data sources), improving accuracy beyond previous possibilities.

5.6 Potential for Exascale and Recommendations

The validation shows that the PROCESS platform scales very well towards exascale and has the potential to fully leverage federated resources and new partners, making even more demanding exascale projects possible. To harness these potentials, it is necessary to continue focused efforts in federalization between computing centres to embiggen the PROCESS computing infrastructure. As mentioned in Chapter 4 it is still necessary to improve specific aspects of the underlying HPC infrastructure of computing centres before true exascale performance can be reached, specifically the improved scalability of the local scheduling infrastructure and support for DTNs to allow moving large amounts of data between compute centres in reasonable time. The validation of the PROCESS components however shows that they are ready to leverage these improvements once they are available.

Besides the infrastructure's potential to reach exascale, we have also shown in this deliverable that each use case achieved immediate benefits through the PROCESS ecosystem while also enabling these use cases potential for further improvements. The generalized approach of PROCESS also provides a great potential to support new use cases and user communities in the future.

Furthermore, a future cooperation and integration with EOSC-related services will enable even broader usage of the PROCESS solution and increase the potential user communities.

6 DEP and market research report

This section gives an overview of the dissemination and engagement plan (DEP) and a summary of the related activities. KPIs updated since the last report are also presented. Finally, the market research report focuses on the potential tier users within the different systems.

6.1 Dissemination and Engagement Plan

In D9.3, an updated strategy was proposed for the final months of the project. It was based on the premise that the incremental nature of PROCESS allows adapting its main components related to the dissemination and engagement strategy. As a consequence, activities were planned to be flexible and customisable in order to adapt communication to the different target stakeholders' profiles.

However, the appearance of COVID-19 and the profound changes that the pandemic has entailed in mobility, and therefore in the possibility of carrying out a large number of activities, has had a great impact on the plans established at the end of 2019.

In March 2020, the consortium established some guidelines to counteract the limitations and ensure in the current situation the correct dissemination of the progress made in the project. It was clear that virtual and online channels were going to be more necessary than ever; as a result, therefore changes had to be implemented in terms of:

- Activities: online conferences, webinars
- Contents: pills, video interviews
- Virtual dissemination materials: extra support materials would be needed instead of roll-ups or other physical elements, such as brochures, infographics

6.1.1 Summary of dissemination activities

A total of 84 initiatives have been implemented in the lifespan of PROCESS. (listed in the annex of D9.5 with a detail of specific activities). More than 60% of them (50) took place in the second part of the project, evenly distributed across its last quarters.

To ensure an effective visibility of PROCESS results and activities, partners have carried out a wide array of different events, offline and online.

Simultaneously, the dissemination of contents has taken place throughout different formats, like: scientific papers (the most common element, with 18 published); blogs, articles and other generic publications. The number dissemination activities by type in RP is displayed in Figure 15.

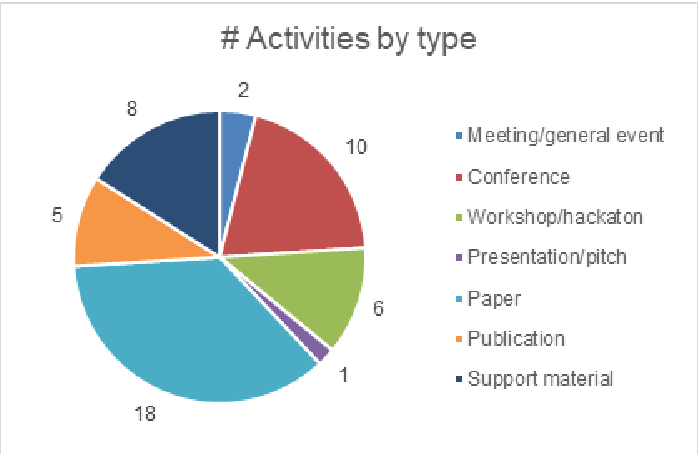

Figure 15: Number of dissemination activities by type, RP2

6.1.2 Specific initiatives for the COVID-19 environment

Infographics

A set of multipurpose infographics[32] were designed to disseminate results of PROCESS through virtual channels. The target audience is the general public potentially interested in the project, not necessarily experts in HPC or in the specific field of activity.

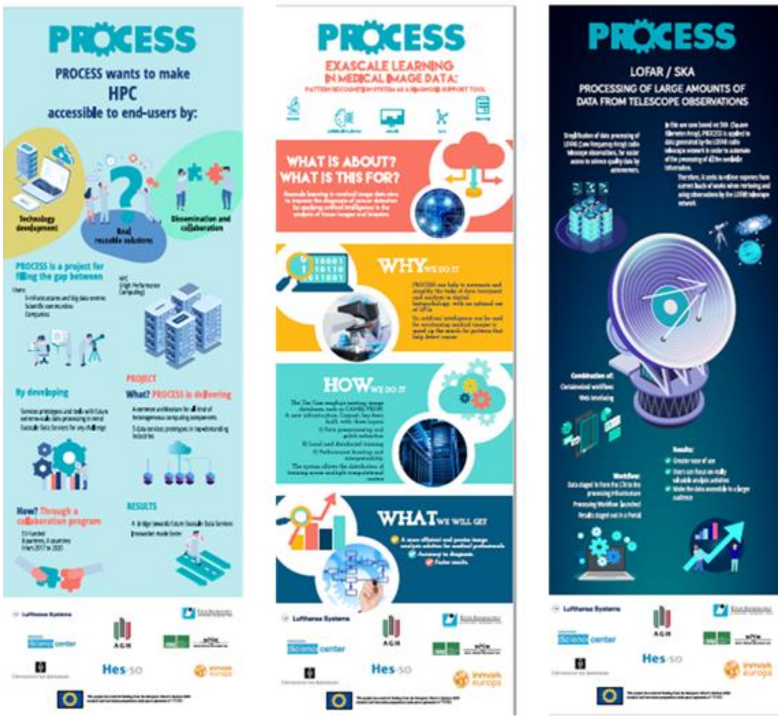

Figure 16: PROCESS infographics available at https://www.process-project.eu/materials/

Videos

Two kinds of short videos were shot to explain in non-technical terms the most important features of the use cases:

- “In a nutshell”: introductory pills of less than one minute.
- “Key facts”: slightly longer recordings explaining the objectives, methods and results of each use case.

They were published on the PROCESS webpage[33], with special coverage through Twitter.

Prototypes

Software developments have received a special coverage from the project’s webpage: apart from a dedicated section as part of the “Downloads” area, the first release of the PROCESS software was disseminated through the news service. The IEE section is displayed in Figure 17.

32 https://www.process-project.eu/materials/

33 https://www.process-project.eu/videos/

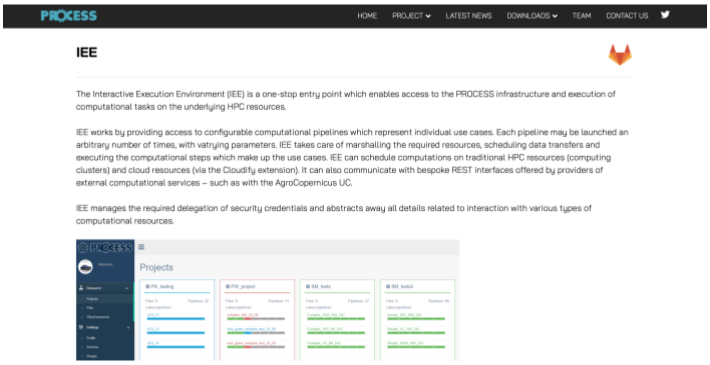

Figure 17: Prototype section at PROCESS webpage

The release was echoed in external media, namely at the specialised publication “Primeur Magazine” (at http://primeurmagazine.com/weekly/AE-PR-10-20-63.html) as shown in Figure 18.

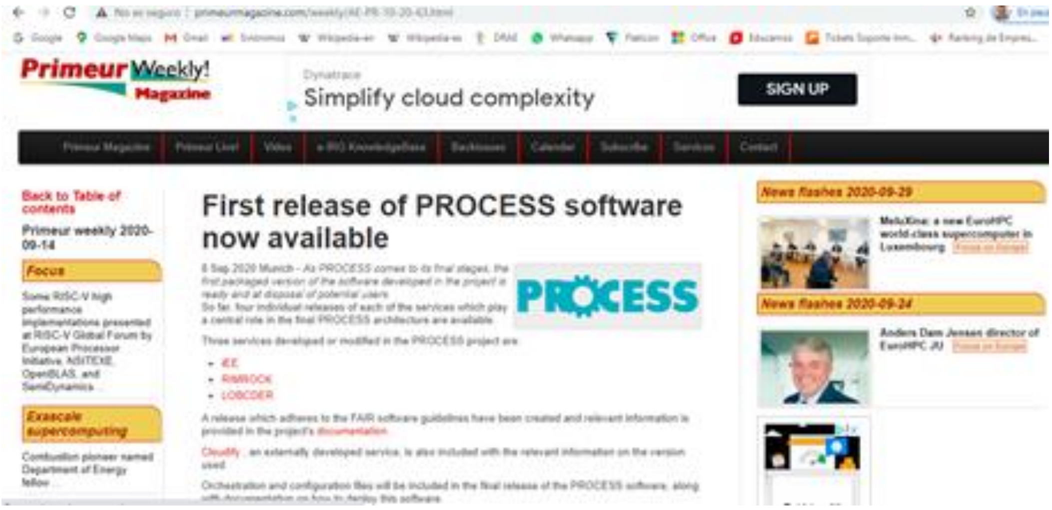

Figure 18: Software release: news at Primeur Magazine

6.1.3 Ongoing dissemination channels

Twitter
Activity of the Twitter account (@PROCESS_H2020) can be assessed through quantitative indicators (KPIs) but it is also interesting to understand the objectives and main activities in this social network as a relevant resource to enhance an active and effective dissemination of the project. The performance KPIS listed in Table 8 show an engagement rate of 0.92 % which is a very good rate for Twitter engagement.

Table 8: Twitter performance KPIs

KPI	Number
Followers	227
Following account	395
Tweets	1019
Impressions per day	56
Engagement rate	0,96%

Earned link clicks	5,39
Earned RTs	6,74
Earned likes	11,97

During the second period of the project, different actions have been carried out in order to improve the power of this tool:

- Tweets have been posted on a more regular basis (see figures below):
  - Publishing "own tweets" on different areas: latest news from the project such as those events where the PROCESS Project has participated; and latest updates in the website (papers, articles, infographics, videos, etc.).
  - "Retweets" from a wide variety of sources (other Twitter accounts), such as PROCESS Project´s partners and key stakeholders: tweets about generic events related to the world of HPC as well as actions to fight against Covid-19.

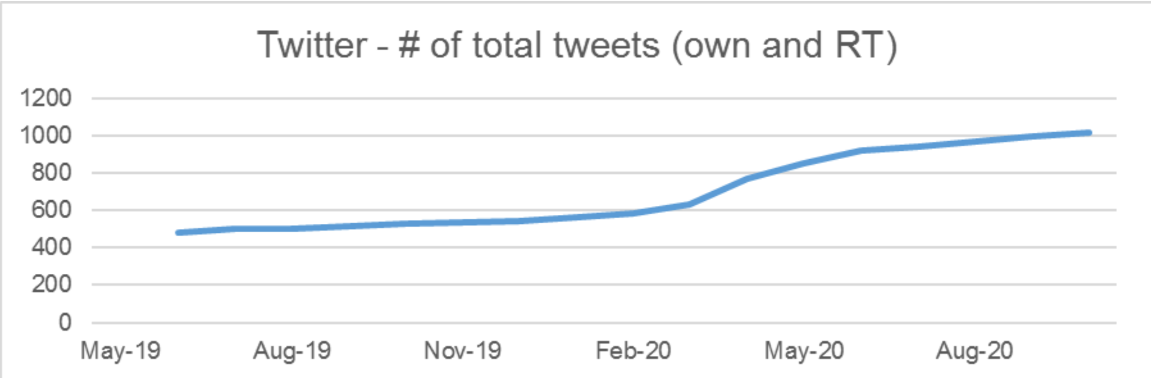

Figure 19: Total tweets, Reporting Period 2

- The account profile has been fine-tuned, in order to improve its efficiency:
  - Bio: the information has been revised according to updated #hashtag policy (see below).
  - Original: PROviding Computing Solutions for ExaScale ChallengeS to allow more intuitive & easier exascale data services. This project receives funding from @EU_H2020
  - New: @EU_H2020 funded project PROviding #Computing Solutions for #ExaScale ChallengeS to allow more intuitive & easier exascale data services - #HPC #BigData.
  - Pinned tweet has been added, with official compulsory info about the project that cannot be included in the Bio (due to the character limitation).

Given the nature and contents of PROCESS, as well as the highly specialised nature of its stakeholders, it was key to keep its audience engaged in Twitter and to foster interaction with the project. Therefore, attention of potential followers has been called in different ways:

- Creating three owned lists associated with the project: Partners (7 members), Key Stakeholders (29 members) and other Stakeholders (94 members), in order to improve both visibility and awareness; also to make it easier for the communication manager to follow them.
- Subscribing to related lists
- Updating the #hashtag policy, to improve their reach capacity. For this, RigeTag tool has been used: for instance, it suggests prioritizing the use of #HPC and #Exascale; using #Supercomputing rather than #Computing or #Supercomputer; #BigData rather than #AI, etc.
- Increasing the level of interactions with both current and potential followers: mainly with likes and RTs on their tweets.
- Posting as many tweets as possible from 1pm to 4pm CET, when more followers are listening. For this, Tweriod tool has been used.

- Expanding the number of followings (accounts we follow) among partners, key stakeholders, companies, institutions and professionals belonging to the HPC world. For this, Audiense tool has been used (it recommends accounts to follow around specific keywords). The current amount of followings is 395, with a relevant increase at the beginning of RP2 (from less than 150 in April-19 to 264 in July-19).

Monitoring Twitter results

A.- Audience:

A.1.- Number of followers

The number has kept on growing steadily along the second and third year of the project, with a slight increase during the last two months. As a result, the threshold of 200 followers was reached in September 2020, with a total of 227 at the end of the project.

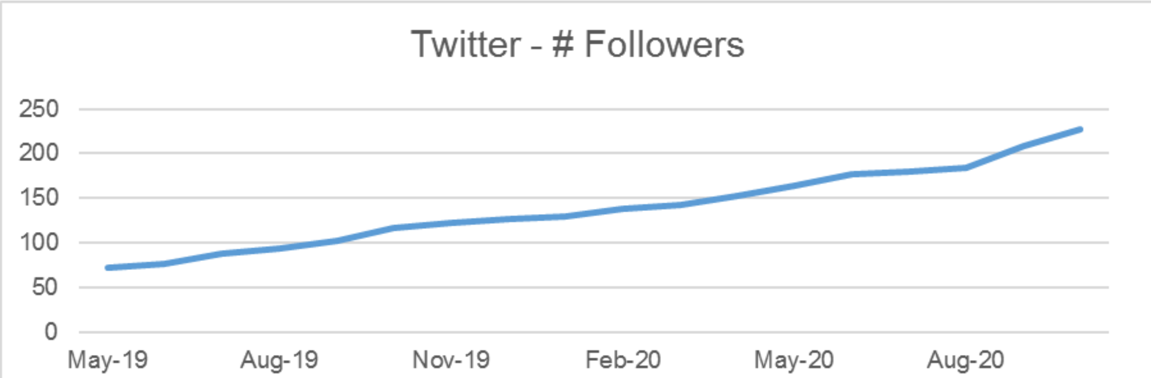

Figure 20: Twitter account: total number of followers

B. Interactions:

B.1 Impressions per day

The average amount of impressions, reached an average of 81 during RP2. Higher levels of impressions are associated with specific moments, like spring of 2020 and the final months of the project.

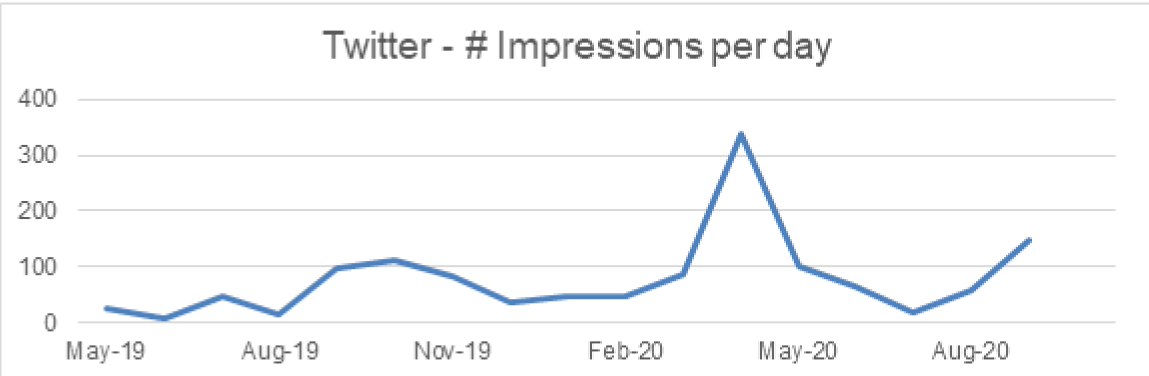

Figure 21: Twitter account: Average number of impressions per day

B.2 Engagement rate

(number of engagements: clicks, retweets, replies, follows, likes, divided by the total number of impressions) It has been 0,96% (monthly average) in this period, with highest peaks in Sep 20 (2,2%), Oct 20 (2,0%), Sep 19 (1,7%) and Aug 20 (1,6%).

Table 9: Twitter account – Engagement rate associated KPIs

KPI	Figure (monthly average)	Peak	2nd
Earned link clicks	5,39	Apr 20 (26)	Oct 20 (25)
Earned RTs	6,74	Apr 20 (32)	Sep 20 (24)
Earned likes	11,97	Apr 20 (68)	Sep 20 (53)

April 2020 was the most active month of all the project, thanks both to a greater number of tweets and an unusually high receptiveness by the audience, followed by Sep 20.

Webpage

The project's web site underwent a series of design and structure improvements at the end of 2019, as explained in D9.3. Thereafter, it has been continuously enhanced thanks to the upload of additional materials, as mentioned before:

- Prototypes
- Videos
- Materials: infographics, as well as brochures and posters
- Architecture section: a new specific page was added to the site in order to explain in more detail PROCESS' architecture and its subsequent advantages

Monitoring Website results

Key metrics referred to the second half of the project (June 2018 – November 2019) show the following activity:

A.- Audience and Behaviour

Main KPIs in terms users and associated behaviour are the following:

A.1.- Visitors: along this period, there were 1121 Internet users who accessed PROCESS webpage.

The period with the most visitors was May 2020 (215) coinciding with the launch of the improved version of the webpage. The last months of the project have witnessed an increase of the number of visitors as well, thanks to the dissemination of the outcomes of the project.

Figure 22: Website: number of visitors per period

Most of the visitors come from Europe (Germany and Spain above all). However, non-European visitors must also be highlighted: 8% from the US, 4% from China and 3% from India.

A.2.- Sessions: there were 2.521 sessions (number of visits to, or interactions with, the website), with 2,32 sessions per visitor on average.

The period when more sessions happened was Oct 2020 (245), with lower but increasing values at the end of the project.

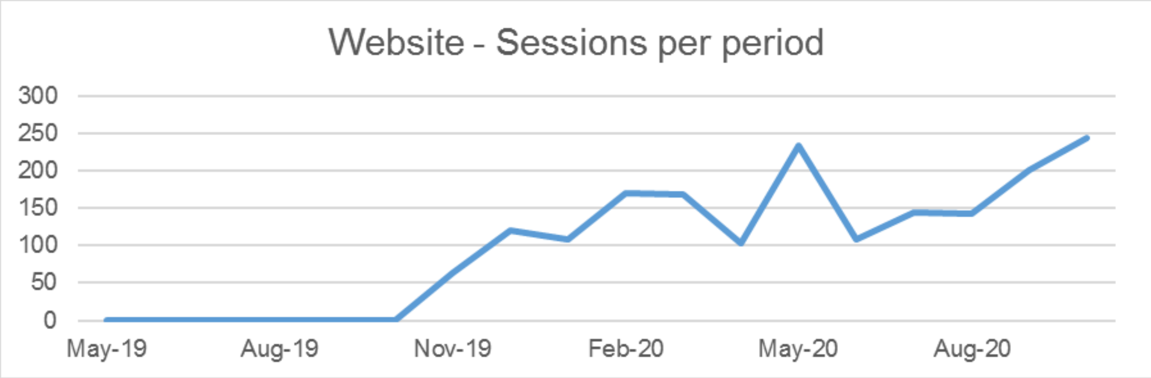

Figure 23: Website: number of sessions per period

As for the quality of the visits and sessions, main KPIs are in the following table:

Table 10: Website KPIs of audience quality

KPI	Value
Pageviews per session (average)	3,26
Length of sessions/ visits (average)	4’ 02”
Bounce rate (percentage of only-one-page sessions – without any further interaction).	49,67%

A.3.- Pageviews: there were 6.140 pageviews (total amount of pages that have been viewed or visited). Of the total pageviews for the period, the homepage was the most viewed (1739), followed by Use Cases (349), Publications (355), and Overview (320).

6.2 Market Research Report

This section is a follow-up of the previous deliverables D9.1, “Initial DEP and market research report“ and D9.3 “Updated DEP and market research report“.

6.2.1 Introduction: a new, unexpected context

Throughout the timespan of the project, an explanatory framework has been used to describe the HPC environment in which PROCESS has taken place (see D9.3 Figure1). As a consequence, market characterisation and segmentation has been developed simultaneously with the analysis of other relevant stakeholders, such as institutional European initiatives or projects related to PROCESS.

Table 11: Main market actors - Key Players in the European HPC Ecosystem

Suppliers: **HPC Centres**	**Potential Users:** **RI infrastructures**	**Partners:** **Agencies and projects**
76 Centres in Europe (HPC500)	1024 institutions (MERIL database)	+50 initiatives
Facilities to install PROCESS software and DM solutions	Beneficiaries of PROCESS developments	Alliances and follow-ups

In conclusion of D 9.3, it was expected that market evolution would follow its rather predictable trajectory and that there would not be necessary to carry out in-depth changes in the analysis, apart from a thorough update.

However, recent changes triggered by the irruption of the COVID-19 worldwide are causing a significant modification of the market structure: market opportunities were existing in 2019 and before, but the sudden alteration of the landscape has distorted their relative weight. This situation advocates for the adoption of a different framework analysis in order to describe and categorise potential segments of users.

Below, we outline the most significant driving forces behind market changes. After that, a subsequent framework will be described in the next paragraphs.

Convergence of HPC with AI, Cloud and Edge/Fog Computing

The following set of technologies are changing the world of computing as a whole. HPC is being also affected by them, both isolated and in combination.

- **AI**’s demands on HPC for boosting calculation capabilities. HPC is becoming a very useful resource for the development of AI; in fact, the combination of HPC and AI is mutually beneficial, with really positive prospects
- **Cloud Computing**, as changing from a pure elastic provisioning of virtual resources (or platforms) to a comprehensive hosting environment. The increasing adoption of Cloud solutions allows access to HPC by new users, and links it to other technologies such as AI and Edge/Fog Computing.
- **Edge Computing** and **Fog Computing**: The booming capabilities of computation and/or storage closer to where it is produced (Edge) and its linkage to Cloud solutions (Fog) pose additional opportunities but also higher requirements for HPC.

Ongoing and expected evolution in HPC infrastructures

As for the capacity and performance of HPC centres, two opposing forces can be traced which shape the offer-demand relationship in the industry:

- Slow-down in the pace of performance growth in the last years:
- Increasing demand due to the traction effect of several technological trends:
    - Artificial Intelligence (AI) + Cloud Computing + Edge/Fog Computing (as explained in the previous point)

  - Containerization in several formats, including highly demanded commercial services
  - GPU-based workloads

As a result, the overall prospects for the market were positive, with an expected CAGR of 7.8% for 2020-2023 at the beginning of that period.

Impact of COVID-19

The pandemic has meant an unparalleled disturbance in every economic activity, and the HPC industry has been no exception. Therefore, favourable market predictions (as shown in the previous paragraph) became obsolete as soon as the impact of COVID-19 was confirmed.

Nevertheless, on the one hand it seems that from July onwards activity is recovering in terms of investments and employment (confirming the hypothesis of "deferral rather than a delay"); on the other hand, from the beginning of the crisis HPC was perceived as a key resource for research to fight COVID-19. These two factors suggest that the crisis will be shorter and less deep than in other fields of activity. In any case, market disturbances are expected to last until at least 2023.

6.2.2 Demand categorisation

As a result of the observed shift in the market trends during the last months, it is possible to draw a more detailed outreach of demand segments, even if potential users of HPC have not changed drastically in the few months.

The guiding force for an updated market framework is the increasing demand for easy-to-use containerised solutions with two axes:

- Support for AI developments
- Interest in specific, industry-focused solutions

Framework

Consequently, a new segmentation model can be traced based on three partially overlapping variables:

- Research Technology Centres ("RIs"): the original target market is still the base of possible future applications and developments of PROCESS' outcomes.
- PROCESS' Field of activity. A set of "1st Tier industries" can be identified, those directly related to the Use Cases, thanks to their technical success and their potential to be applied through the activity.
  - UC#1: Health
  - UC#2: Astronomy
  - UC#3*:* Multipurpose platform
  - UC#4: Airlines
  - UC#5: Agriculture and environment analysis projects
- Organisations looking for Artificial Intelligence solutions: companies or institutions becoming "AI seekers" in an expanding market for all kinds of related services.

Figure 24 shows the three segmentation variables and their possible intersections:

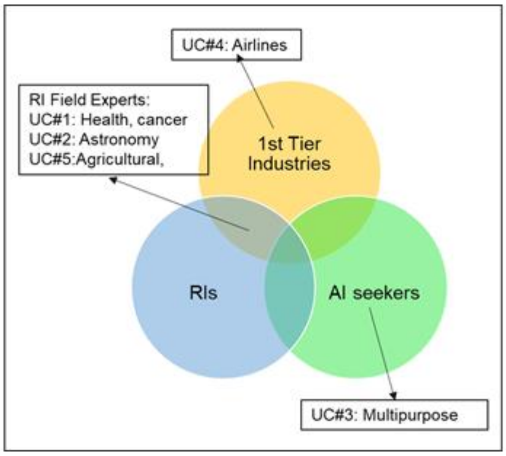

Figure 24: Segmentation variables and possible segments of interest

Therefore, priority subsegments can be identified whether in the intersections of the variables, or in separate areas of the diagram.

Description and quantification;

The annex in D9.5 contains a list of the identified members of main users' segments, which can be described as follows:

First-Tier potential users (134 organisations):
RI in fields of activity - Research Infrastructures related to 1st Tier Industries are arguably the most interesting potential users of PROCESS outcomes:

- UC#1: Health and cancer research institutions, with focus on imaging and exploration activities
- UC#2: Astronomical centres and related infrastructures and offices
- UC#5: Earth observation and agricultural research agencies.

Altogether, these three sub segments account for a considerable share of total RI infrastructures in Europe. Initially, 109 Centres have been identified whose main activity is directly related to Use Case #1 (51 RIs), UC#2 (41 RIs, including LOFAR and SKA themselves) and #5 (17 RIs).

They account for roughly 10% of the 1.042 Centres related to Europe (as in MERIL database from Sep-2020), with a high share (one third) in the domain of "Physics, Astronomy, Astrophysics and Mathematics".

Table 12: RI centres by most important domain

Domain	UC #1	UC #2	UC #5	None	Total
Biological and medical sciences	49	-	3	291	343
Physics, Astronomy, Astrophysics and Mathematics	-	36	-	72	108
Earth and Environmental Sciences	-	3	10	141	154
Chemistry and Material Sciences	-	-	2	113	115
Information Science and Technology	-	2	-	35	37
Not declared	2	-	2	105	109
Engineering and Energy	-	-	-	51	51
Humanities and Arts	-	-	-	85	85

Social Sciences	-	-	-	40	40
Total	**51**	**41**	**17**	**933**	**1042**

The potential user RIs are widely scattered among European countries, having identified institutions from 26 different countries.

Table 13: RI centres by country

Domain	UC #1	UC #2	UC #5	None	Total
France	5	7	4	89	105
Germany	4	6	2	109	121
Spain	2	7	1	56	66
Netherlands	5	2	2	76	85
Finland	6	1	1	25	33
United Kingdom	-	3	3	72	78
Ireland	6	-	-	30	36
Italy	-	6	-	29	35
Hungary	3	1	1	46	51
Switzerland	3	1	-	33	37
Other countries	17	7	3	368	395
Total	**51**	**41**	**17**	**933**	**1042**

Top-priority airlines - non-competitors

There is a clear opportunity to use outcomes from UC#4 by airline companies non-concurrent with Lufthansa Group. That means either airlines that operate in different markets or companies which are already partners of Lufthansa Group in other initiatives. That leads to the Star Alliance partnership, from whom Lufthansa is a founding member. It also constitutes the largest partner alliance (25 full members apart from Lufthansa itself), so it can be reasonably assessed as PROCESS' main potential market for exploitation of the results of UC#4.

Second-tier potential users (127 organisations):

- **European health/cancer leading researchers**
  According to Nature's Index 2020 in Cancer Research, there are European 46 institutions (including LMU) as part of the worldwide top 200.
- **Astronomy**
  - 60 main potential users have been identified. They are formed by members of SKA and LOFAR, national representatives of the IAU (International Astronomy Union), members of the European consortium Astronet (https://www.astronet-eu.org/) and main national astronomy associations.
- **Agriculture/environment**
  - Two groups of institutions, comprising 21 are considered of high potential for using the outcomes of PROCESS: partners of the EIONET ("European Environment Information and Observation Network") consortium and members of Copernicus' Action Group on Land Monitoring ("EAGLE Group").

Third-tier potential users:

- **Other Research Infrastructures in Europe**
  As seen in the First Tier, centres with activities non-related to PROCESS Use Cases represent more than 900 potential users across Europe.

- **AI seekers**
  Demand for AI in Europe is expected to climb up to 22.000 M€ in 2021[1], with an increase of its share of the total world market to 18%, from 15% in 2025. to 18%, € in 2025 (18% of the worldwide market), as European industries are beginning to adopt intensively AI tools and applications

It is important to consider that pervasiveness of AI means that only a small fraction of all potential demand (less than 15%) belongs to sectors linked to PROCESS' Use Case. Nevertheless, the expected market could amount to 1100 M€ in Government, Transportation and Education organisations in Europe in 2021.

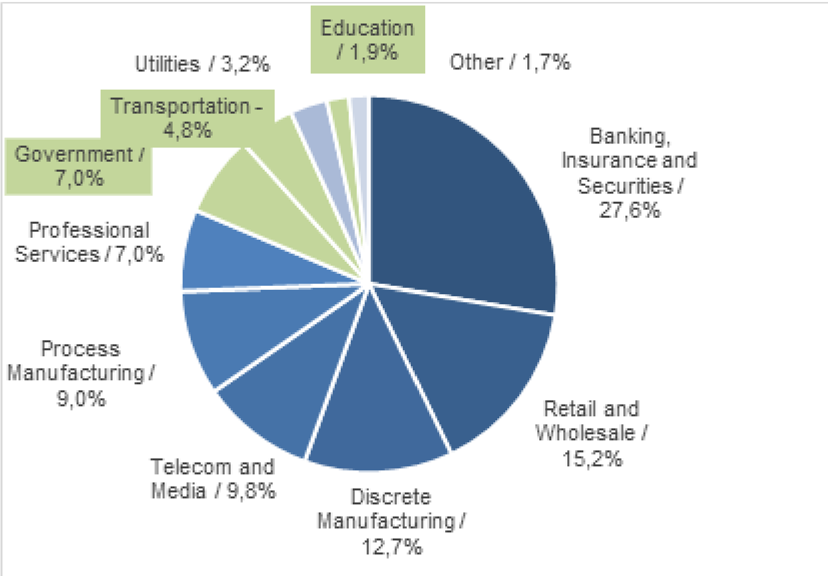

Figure 25: Potential demand of AI in Europe by industries in 2021, Adapted from IDC[34]

[34] IDC Worldwide AI Spending Guide, Forecast, August 2020 and IDC Worldwide Robotics Spending Guide

7 Conclusion

This deliverable presents the final report of the PROCESS project's Use Case prototypes, the reference exascale architecture and the derived PROCESS architecture as well as the evaluation and validation of the platform. While most services and prototypes have been almost final in previous deliverables, D1.3 presents the final updates and summarises the final outcomes and outputs of this project.

We show that each use case benefits from the PROCESS ecosystem and that the modularity of the PROCESS building blocks makes it very easy for each use case to focus on the most important features, showing the ecosystem's focus on adaptability. Each use case prototype was released publicly within the PROCESS software release and can easily be adapted and reused be new users.

Since the goal of PROCESS was to show the potential of exascale computing right now, one of the most important aspects of the project's outcomes is the reference exascale architecture and from that our derived PROCESS architecture. Since there are still many steps to take to build true exascale systems, approaches like the PROCESS architecture are necessary to bridge the gap between the present HPC landscape and the exascale future we strive to achieve. Our in-depth evaluation and validation of the architecture, as well as the platform performance through performance modelling shows its potential to scale towards exascale. We also show that the PROCESS ecosystem is ready to incorporate future innovations and to grow its potential further through that.

While we show great potential within the Use Cases and the PROCESS platform for scalability and support of different user communities, it is very important to reach these new user communities. The project's focus therefore has always been on reaching new users and communities to further enhance the ecosystem by learning their requirements but also to give those users and communities the chance to learn about the potential that PROCESS may offer and to easily use our ecosystem. We therefore conclude this deliverable in the chapter 6 where we give a detailed overview of the project's dissemination activities, making sure that the projects results will live on in the future, beyond the scope of the project and further enhance new use cases and communities.