

Data Pitch

H2020-ICT-2016-1

Project number: 732506

D2.4 Updated EaaS and support

**Coordinator: Stefano Modafferi (University of Southampton IT
Innovation Centre)**

Quality reviewer: Jeremy Decis, Dawex

Deliverable nature:	Other
Dissemination level: (Confidentiality)	Public
Work package	WP2
Contractual delivery date:	31/12/2018
Actual delivery date:	02/01/2019
Version:	V1.0
Keywords:	Infrastructure

Table of Contents

Abstract	3
Executive summary	4
1 Introduction	5
2 Experiment as a Service (EAAS)	6
2.1 Hosting services	6
2.2 Pre-Processing Services	6
2.3 Data access verification	7
2.4 Technology consultancy services	7
2.5 Accounting Services	7
3 Conclusion	8

ABSTRACT

Following the Everything-as-a-Service (XaaS) approach Data Pitch implements the Experiment as a Service (EaaS). EaaS is based on the idea that the experimenters have to be able to perform their experiments in a seamless way regardless of geographic or organisational separation of provider and consumer. The way in which Data Pitch supports the paradigm varies from case to case and mostly is about removing all the technological and operational barriers that may be existing in the performing of an experiment.

This report is an updated version of D2.2 and describes the different technological and infrastructure services offered by Data Pitch to experimenters and Data Providers.

They are divided into: hosting services, pre-processing services, data access verification services, technology consultancy services and accounting services.

All the services are activated upon request. The current engagement model foresees that WP5 (Competitive call) is the front-end toward the experimenters and WP3 (Data and Data Providers' liaison) is the front-end with Data Providers. WP2 (Data experimentation facilities) can be engaged at short notice by any of the other work packages if a need for any of the services arises.

EXECUTIVE SUMMARY

This short document is part of the D2.4 deliverable that is classified as “other”. This means that the report describes the services provided by Data Pitch to the SMEs for supporting the experiments. They are explained in Section 2 where they are divided into: hosting services, pre-processing services, data access verification services, technology consultancy services and accounting services.

D2.4 is an updated version of D2.2.

1 INTRODUCTION

This deliverable describes what type of support Data Pitch offers to the SMEs for the execution of their work plan mainly from a technological and infrastructural point of view.

Data Pitch supports the execution of the work plans without imposing any specific type of solution. In fact, Data Pitch offers a set of services to both the Data Providers and the SMEs selected for the acceleration period (hereafter also called experimenters). A constant dialogue between the Data Pitch team and the counterparts ensures the SMEs and the Data Providers know what services are available and they will pick the set of services best suitable for their case.

The services offered by WP2 complement the ones implemented in WP5. In fact, the support provided by WP2 ranges from preliminary services to enhance or clean data, to the actual capacity of hosting experiments in Data Pitch infrastructure. WP2 also when possible and appropriate (e.g. when an API for accessing the data is used) offers accounting service to monitor the access and usage to the data independently. Finally, the persons involved in WP2 also act as technical experts to advise the SMEs in the setup of the work plan.

Experiments are characterized by a high level of innovation that comes with a similar level of uncertainty, so flexibility must be ensured. The support model WP2 adopts to support the required flexibility is the “management by exception”. Besides a constant involvement in the preparation and the execution of the work-plan in the acceleration period, WP2 is ready to promptly react to uncertainty and unexpected technical requirements and challenges. In fact, as also the first round of acceleration has confirmed, the cutting-edge nature of the proposed solutions and the limited mutual knowledge of Data Providers and SMEs could require to revisit the approach to solve the problem. The front-end of the interaction with the SMEs remains WP5 that engages with WP2 to cover the technological aspects.

2 EXPERIMENT AS A SERVICE (EAAS)

The experimental platform offered by Data Pitch as described in D2.3 is available for the first cohort of the experiment. This section describes the available services.

2.1 HOSTING SERVICES

From a hosting point of view as described in D2.3 there are several possibilities:

- Data can be hosted in the data provider infrastructure.
- Data can be hosted in the experimenter infrastructure.
- Data can be hosted in a commercial cloud service.
- Data can be hosted in one of the Data Pitch infrastructures

Each of the above solutions addresses different requirements and scenarios and the final decision of where to host the data is taken as a joint decision by the consortium members, the data provider, and the experimenter.

The second cohort is currently undergoing the negotiation phase in which the fine-tuning of the experiment details are decided. It is envisaged that hosting two data sets (Bloomberg and UK Met Office). This is still not finalised and might change.

The Bloomberg dataset is characterized by a large number of PDF files that need to be parsed and analysed using NLP techniques. The dataset contains only historical data and it is not going to change during the experimentation period.

The UK Met Office offers forecast prediction and it is a live stream (frequency every day). Automatic synchronization of the master source (the Met Office) and the hosting facility will be implemented.

2.2 PRE-PROCESSING SERVICES

Data Pitch offers to the Data Providers preliminary services to be implemented on the dataset object of the experiments. The primary type of this class of service is the possibility of anonymising the data. Various techniques are supported. The methods will be chosen according to the different actual scenarios and they can be divided into¹:

- Non-perturbative methods (that reduce the detail in the data by generalization or suppression of certain values (i.e., masking) without distorting the data structure.
- Perturbative methods do not suppress values in the dataset but perturb (i.e., alter) values to limit disclosure risk by creating uncertainty around the true values.

The interested reader can also be referred to the paper of Hunderpool² for a discussion on the best techniques to preserve anonymity still maintaining statistical value. As clearly explained in all the literature, there is never a one size fits all solutions and each case requires an analysis to identify the best instruments. This is precisely the approach Data Pitch is following.

Other types of services include for instance a light data cleaning if this is not already a task identified in the work plan of the experiment.

So far no Data Providers have required services to be implemented on their data; in fact, very often they have solutions themselves.

¹ https://sdcppractice.readthedocs.io/en/latest/anon_methods.html#hdfg12

² Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E.S., Spicer, K. and De Wolf, P.P., 2012. *Statistical disclosure control*. John Wiley & Sons.

2.3 DATA ACCESS VERIFICATION

Another type of service supporting the experimentation is the verification of the access to the data. WP2 is in the process of verifying that the data to be used in the experiments are effectively what promised and described in the challenge and that they are actually accessible. The Data Providers produce the final list of the dataset involved in the acceleration period. The datasets are uniquely identified. A WP2 member does a one-off test that the data are actually accessible.

This service is important also from a formal point of view because it is required to Data Pitch to remove all the barriers to the execution of the experiments as soon as possible. A live registry of all the used datasets has been created as part of the process implemented in the second cohort. This registry was not present in the current, easy to access, form in the first cohort.

2.4 TECHNOLOGY CONSULTANCY SERVICES

The WP2 team has been contributing with technology expertise to the Data Pitch life since the inception of the project. This paragraph is specifically about the services activated by WP5 team when specific expertise on the technology aspects is required during the negotiation phase and the acceleration execution period.

Besides the general availability to support the technology-related topics, WP2 is specifically involved in discussing with experimenters and providers during the work plan to:

- Support the SME in defining the technical challenges they expect during the experimentation period.
- Support the SMEs in defining specific, measurable, agreed upon, realistic and time-based (SMART) technical objectives.
- To align and clarify from a technical point of view the expectations between the providers and the experimenters on the outcome of an experiment.

2.5 ACCOUNTING SERVICES

When offering hosting services, Data Pitch also includes accounting services for checking what type of data and how many types of data are accessed by the experimenter.

The other case in which Data Pitch offers accounting service is when data are hosted in the Data Providers' infrastructure and the access to them is made through API. In this case, a proxy is implemented, such that the request from the applicants is forwarded to the providers and the results pushed back to the applicant. This is an optional service offered by Data Pitch. This service serves two purposes: it offers secure access to only the data agreed to be shared and it checks the amount of data exchanged. Simple metrics are then extracted and used to understand how many time and how much data are accessed during the experimentation period.

3 CONCLUSION

This document discussed the different solutions available in Data Pitch to support the experiment as a service paradigm. The variety of scenarios and requirements that the challenges present require a flexible approach and an interaction with all the involved parties to identify the best solution. For this reason, besides the constant interaction already planned, a “management by exception” approach is also implemented. WP5 is identified as the front-end in the interaction with the SMEs, and WP2 team can be engaged at very short notice to support the experiment providing hosting capabilities and/or consultancy service as required.