Data Pitch

H2020-ICT-2016-1

Project number: 732506

D2.2 EaaS facilities and support

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

Quality reviewer: Johanna Walker,

Web and Internet Science, University of Southampton

Deliverable nature:	Other
Dissemination level: (Confidentiality)	Public
Document URL	https://drive.google.com/drive/folders/0B2SsjlchXofDVEhFYUFlS1dmMWs
Work package	WP2
Contractual delivery date:	31 December 2017
Actual delivery date:	11 January 2018
Version:	V1.0
Keywords:	Infrastructure

Table of Contents

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 organizational 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 existing in the performing of an experiment.

This report 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 with 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 noticed 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 D2.2 deliverable that is classified as "other". This means that the report complements the services provided for the experiments.

The document describes the technology and infrastructure related services offered by Data Pitch to experimenters and data providers.

1 INTRODUCTION

This deliverable describes what type of support Data Pitch offers to the experiment execution mainly from a technological and infrastructural point of view.

Data Pitch supports the execution of experiments 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 they know what services are available, but, possibly after consulting with Data Pitch, they will pick the set of services best suitable for their case.

The services offered complement the ones implemented in WP5. In fact, the support offered 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 independently monitor the access and usage to the data. Finally, the persons involved in WP2 also act as experts in the discussions with the selected SMEs when fine-tuning of the technology aspects of the agreed experiment is necessary.

At the time of submitting the deliverable the experiments are not yet started (scheduled for 1st Feb 2018 – 31 July 2018).

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". As soon as something is necessary a prompt, quick reaction is to be implemented. The front-end of the interaction with the SMEs remains WP5 that engages with WP2 when specific technology or infrastructure service/consultancy are necessary.

2   EXPERIMENT AS A SERVICE (EAAS)

The experimental platform offered by Data Pitch as described in D2.1 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 D2.1 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
    - IRIDIS infrastructure
    - IT Innovation infrastructure
    - Secure lab infrastructure

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 first of cohort of experiments focuses on historical data (i.e. no live data are used) and the preferred method of hosting is the use of the experimenter's infrastructure. So far no one has yet required the hosting of the data by Data Pitch.

If problems should arise and/or the situation requires Data Pitch to host the data (not all the experimental scenarios are already fully defined), WP2 is already able to support the scenario.

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 main type of this class of service is the possibility of anonymising the data. Various techniques can be implemented and are available according to the different possible scenarios. 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. Often they have solutions themselves and in a specific case, the provider decided to commit some effort to create a suitable tool that he will be able to reuse in the future.

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 involve in the acceleration period. The dataset 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 for the execution of the experiments as soon as possible.

2.4   TECHNOLOGY CONSULTANCY SERVICES

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 definition when:

- A technology validation and/or a refinement is needed to make clear and accountable what is going to be developed during the acceleration period
- When there are different expectations between the providers and the experimenters on the outcome of an experiment. In this case, a clear understanding of the technology challenge is important to solve the issue and WP2 supports it.
- When the technology soundness of the detailed proposed solution is necessary (the application was based on a higher level description)

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 a 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, a "management by exception" approach is implemented. WP5 is identified as the front-end in the interaction with the SMEs, and the involved persons know that WP2 can be engaged at very short notice to support the experiment providing hosting capabilities and/or consultancy service as required.