

Data Pitch

H2020-ICT-2016-1

Project number: 732506

D6.4 Training Materials, Learning Curriculum and Webinars

Coordinator: Johanna Walker, University of Southampton

With contributions from: Huw Fryer, University of Southampton; Jeremy Decis, Dawex; Deborah Okenla, ODI.

Quality reviewer: Heidi Lindvall, ODI

Deliverable nature:	REPORT
Dissemination level: (Confidentiality)	PUBLIC (PU)
Nature	OTHER
Document URL	
Work package	WP6: Dissemination, community building, communication
Contractual delivery date:	30th June 2017

Actual delivery date:	30th June 2017
Version:	1
Keywords:	training, data science

Table of Contents

Abstract	3
Executive summary	4
Introduction	5
Current Training Materials	6
Data Science and Skills	6
Open Data	8
Digital Business Skills	9
Business Innovation	11
Legal and Privacy Aspects of Data	12
Applicant Training and Guides	13
Planned Training Materials	14
Annexes	15
A: ODI Courses	
B: EDSA ODI Course	
C: Pseudonymisation Power Point Presentation	

Abstract

This report discusses the training, learning and support materials and curriculum compiled and developed for Data Pitch audiences.

Executive summary

This report outlines the purpose of and audience for training within the Data Pitch ecosystem. It details courses and materials available in 6 key areas; data science and skills; open data; skills for digital businesses, in areas as diverse as marketing and cyber-security; business innovation, the legal aspects of sharing data and training on the application process.

Introduction

In this task we develop a curriculum for data-driven innovation for three types of audiences: (i) companies like the ones we will incubate, interested in new business models, products, services, and networks, enabled by data and a participatory approach to innovation; (ii) industry in general that is considering an open innovation lab around their own data assets; and (iii) open innovation researchers and practitioners.

It will include a part that is related directly to Data Pitch, covering the materials developed in WP2 around the use of the experimentation facilities, but also, more importantly, general information about data science methods, data literacy, data privacy, data sharing and business models, entrepreneurship, and established technology stacks. Some of these materials will be developed in the project, including the user guides and tutorials presenting our data and computing infrastructure in WPs 2 and 3. We will focus on Web-based training and will offer two live webinars for each round of the call focusing on technology topics, recordings of which will be made available to download through the project's online communication channels. We will build upon a rich collection of learning materials available at SOTON around data science topics complemented by ODI's programme around data-driven innovation, and will align these efforts with the extensive work on privacy developed in the project.

1. Current training materials

1.1 Data science and skills Courses

Developing data science and skills are at the centre of the Data Pitch aim to support a healthy data-driven, open innovation ecosystem.

1.1.1 Massive Open Online Courses

Finding Stories in Data

Finding Stories in Data takes learners through the process of developing a compelling data narrative. Skills covered include visualising data; finding and telling stories in data; gathering data and cleaning data.

This is delivered online by the ODI and is available through the European Data Science Academy. It is a free course.

More detail is provided in Annex B.

Linked Data and Semantic Web

This course guides learners through the basics of Linked Data and the Semantic Web - exploring how this new Web of Data isn't about creating a big collection of standalone datasets, but is instead about using a common format to ensure data is interrelated. As well as the background to the semantic web, learners are introduced to linked data technologies and engage in writing queries in SPARQL.

This is delivered by the University of Southampton and is available through the European Data Science Academy and FutureLearn.

<https://www.futurelearn.com/courses/linked-data>

1.1.2 Online courses

Fundamentals of Data Science

In this course, learners understand the foundations of the data science process; are taught to evaluate data science tools and techniques based on their suitability for particular tasks; and gain hands-on experience using R to analyse data. The course emphasises a hands-on approach to learning data skills, offering a number of interactive, online exercises to try out many of the techniques and concepts covered in the taught material. This is aimed at analysts and technical teams, and there is a complementary non-technical version for managers (see below).

This is delivered by the University of Southampton and is available through the European Data Science Academy.

<http://courses.edsa-project.eu/course/view.php?id=48>

1.1.3 Written materials

Foundations of Data Science eBook

This interactive book supports self-directed study and covers both the architecture and analytics aspects of data science.

This is delivered by the University of Southampton and is available through the European Data Science Academy. It is a free book.

<http://courses.edsa-project.eu/mod/page/view.php?id=299>

1.2 Open data

Participants in Data Pitch are encouraged to support their use of closed and shared data sets with open data, and these courses will develop their knowledge in this area.

1.2.1 Workshops

Open Data in a Day

Open Data in a Day is a one-day interactive course run by experts. It has 4 sections: Discovering open data; Copyright, licensing and open data; Making data usable; and Getting hands-on with open data.

This is delivered by the Open Data Institute and is a paid course. More detail is provided in Annex A.

Open Data in Practice

A three-day interactive course for those who have experience with open data and have knowledge of the definitions of open data and best practices for publishing open data. The course covers the very best practice in using and publishing open data and the legal and policy requirements in order to remove any potential risks.

This is delivered by the Open Data Institute and is a paid course. More detail is provided in Annex A.

1.3 Digital business skills

This strand of training covers skills and technologies for data businesses.

1.3.1 Massive Open Online Courses

Digital Marketing: Challenges and Insights

This course focuses on emerging trends in digital culture and online consumer behaviour, data analytics and privacy. Throughout the course, the implications of these developments for both marketers and consumers is examined. Week 1 examines the evolving changes in how consumers behave online and what this means for the organisations they interact with; Week 2 discusses developments in technologies we use and the relationship between “online” and “offline” and Week 3 examines the opportunities and challenges in obtaining value from the vast amounts of social data now generated, at a time when much uncertainty still exists about personal privacy and control of online assets.

This course is delivered by the University of Southampton and available through FutureLearn.

<https://www.futurelearn.com/courses/digital-marketing>

Secure Android App Development

It is expected that a number of applications and successful proposals will suggest an app as the final product. This course provides an introduction to mobile app security, including common vulnerabilities found in Android apps, and how to detect and mitigate them, as well as how to provide proof of secure code. Architecture, permissions, interprocess communication and static analysis of code will be covered as well as the use of a popular toolset to identify and fix vulnerabilities.

This course is delivered by the University of Southampton and available through FutureLearn.

<https://www.futurelearn.com/courses/secure-android-app-development>

1.3.2 Online courses

Data Science for Digital Marketing

The course is aimed at management rather than technical or analytic teams. It emphasises a hands-on approach to learning the data skills needed by business managers and executives, covering a number of techniques and concepts.

This course is delivered by the University of Southampton through the Southampton Data Science Academy. It will launch in autumn 2017.

1.3.3 Webinars

Monetizing your data

This webinar that guides potential data sharers through the process of sharing their data for commercial purposes.

The webinar is created by Dawex and is available from their web site.

<https://www.dawex.com/members/webinar-monetizing-data>

1.4 Business innovation

Our business innovation offering looks at the interaction between data and organisations from a strategic point of view.

1.4.1 Workshop

The Business Innovation workshop is designed to help companies unlock the value of open data in business. The course enables participants to: Analyse existing open data business models; Build a business model for an open data focused business; Present a strategy for leveraging open data, and, Identify future opportunities for business and sector innovation.

This course is delivered by the Open Data Institute. It is a paid course. More detail is provided in Annex A.

1.5 Legal and privacy matters pertaining to data sharing and use

Although legal and privacy matters are also touched on elsewhere, these materials provide clear guides to the frameworks and regulations around data sharing and use.

1.5.1 Written materials

Legal and Privacy Toolkit

The legal and privacy toolkit is a crucial component of Data Pitch. It guides users through the question - "What are the critical things to consider when sharing and reusing data for defined innovative purposes under the Project?" Its focus is on ensuring that legal rights in relation to shared data are respected, and legal obligations in respect of such data followed. In particular, information is provided on how organisations should anonymise personal data before it is shared for analysis, and ways to mitigate the risk of harm befalling data subjects as a consequence of data relating to them being processed in new ways (for secondary purposes).

The legal and privacy toolkit aims to provide an overview of the legal and regulatory framework that applies to data sharing and data reuse, especially in respect of privacy and data protection law obligations. It also sets out key considerations governing the data sharing arrangements underpinning this Project in order to ensure that sufficient safeguards are in place to address legal issues that might arise, such as in respect of licensing intellectual property rights.

The toolkit achieves this overview with reference to existing research and guidance that reflect best practices in this field (including from guidelines, checklists, opinions, and recommendations) based on EU standards. To this end, the toolkit takes into consideration relevant features from different legal systems across the EU that restrict the sharing and reuse of datasets, together with associated rights that are raised. It also set out the outline of a methodology to guide participants in the Data Pitch project from different countries through this area in a way that implements a uniform mechanism.

This toolkit will be updated as necessary and a final version published in 2019. It is available on the datapitch.eu website.

Pseudonymisation Guide

We have developed a brief presentation for data owners considering sharing their data, regarding the process of data pseudonymisation, and how to facilitate compliance with the General Data Protection Regulations coming into force in May 2018.

This presentation is available in Annex C.

1.6 Training materials supporting applications to Data Pitch

In order to facilitate the highest standard of application materials will be provided to support applicants and provide a guide to the application process.

1.6.1 Webinars

Catalogue and submission platform tutorials

Tutorials are being developed that will guide participants through the use of our data catalogue and the submission platform. A demonstration of the submission platform can be found at <https://www.f6s.com/datapitchaccelerator/apply>.

Call webinar

As well as these tutorials a more detailed webinar is being developed that guides potential applicants through the application process, from selecting a challenge, to understanding the criteria for assessment to ensure a complete application form, to the timings of the process. This will be broadcast and then made available for download from the Data Pitch site.

2. Planned training materials

As well as developing new materials for the acceleration part of our activities, we will also be reviewing our offer for the second call, which will launch in July 2018.

2.1 Massive Open Online Courses

Computational Thinking

In Month 10 the University of Southampton will launch a MOOC on Computational Thinking. the thought processes involved in formulating a problem and expressing its solution(s) in such a way that a computer—human or machine—can effectively carry out.

This will be delivered by the University of Southampton and will be available through the European Data Science Academy. It is a free course.

2.2 Written materials

Open innovation accelerator materials

In Month 13 successful applicants from the first round will enter the accelerator. Starting in July we will develop materials for the start ups to ensure they can access the appropriate data or management skills.

We will also develop materials about the accelerator experience, particularly those of interest to open innovation researchers and practitioners.

Annex A: Open Data Institute training courses

Open Data In A Day

A one-day interactive course run by our experts. You will learn how to discover, use and describe the benefits of open data, and how they can impact the physical activity sector. This course costs £269 +VAT for an earlybird ticket or £299 +VAT for a full price ticket. [Open Data in a day](#) course scheduled for the [21st June](#) and the [31st of July](#).

There are four sessions during the day:

Session 1: Discovering open data

Session 2: Copyright, licensing and open data

Session 3: Making data usable

Session 4: Getting hands-on with open data

Each session runs for between 1.5 and 2 hours, with lunch 12:45 - 1:45. Those who attended the previous training in December found it incredibly insightful - "The ODI are ushering in a new age of Open Data and their Data in a Day course is a vital resource for any business that is serious about staying ahead of the curve." Please follow this hyperlink for [more information](#) on Open Data in a day.

Open Data in Practice

A three-day interactive course for those who have experience with open data and have knowledge of the definitions of open data and best practices for publishing open data. It is for those who are looking to develop the necessary skills to take advantage of open data in their organisation. The course will cover the very best practice in using and publishing open data and the legal and policy requirements in order to remove any potential risks.

The course focuses on three key areas (1 per day)

Day 1: Publishing open data

Day 2: Exploiting open data

Day 3: Analysing and visualising open data

Please follow this hyperlink for [more information](#) on Open Data in Practice. This three day course costs £999 (ex-vat) or £699 (ex-vat) for our members. To sign up for the training [please click here to sign up](#).

We recommend taking [Open Data in a Day](#) prior to attending, if you are new to this area.

Business Innovation

This workshop is designed to help you unlock the value of open data in business. This course costs £249 (ex-vat) or £169 (ex-vat) for our members.

Over the course of the day, you will discover:

- Key examples of businesses innovating with open data
- Opportunities and challenges open data presents for businesses
- How to take advantage of open data in your business

- How to build a better relationship with 'open'
- The effect of opens on commercial relationships

By the end of the course you will be able to:

- Analyse existing open data business models
- Build a business model for an open data focussed business
- Present a strategy for leveraging open data in your business
- Identify future opportunities for business and sector innovation

The workshop is suitable for any startup, entrepreneur or established business developing open data products or services. We have just arranged this workshop for Tuesday 25th July, 2017. If you wish to sign up for the training [please click here to sign up](#).

Annex B

The European Data Science Academy (EDSA) designs curricula for data science training and data science education across the European Union (EU). EDSA establishes a virtuous learning production cycle whereby we: a) analyse the required sector specific skillsets for data scientists across the main industrial sectors in Europe; b) develop modular and adaptable data science curricula to meet industry expectations; and c) deliver data science training supported by multi-platform and multilingual learning resources. The curricula and learning resources are continuously evaluated by pedagogical and data science experts during both development and deployment. ODI (e-learning): *When effectively analysed and presented in a clear and compelling way, data has the potential to create impact. Whether that's changing perceptions, offering counterintuitive insights or prompting action, impact happens when data acts as the catalyst for change. And at the heart of driving change is the skill of finding and telling stories using, where relevant, compelling visualisations. No field is more experienced at finding and telling stories than journalism, and no field better at using data than data science. This set of modules looks at what these fields can learn from each other, in order to find and tell compelling stories with data. This is a free course, i.e. there are no fees for registering and attending it.* Courses available [here](#)

1. Introduction to data storytelling
2. The four step process
3. Understanding your rights to use data
4. Gathering data
5. Organising data
6. Cleaning data
7. Filtering and pivot tables
8. Data visualisation formats
9. Data visualisation best practice
10. Visual deception
11. Narrating your story

Annex C: Pseudonymisation Power Point

Sharing 'pseudonymised' data with SME winners of the Data Pitch (EU Horizon 2020) competition

- Data that has undergone pseudonymisation (through the removal or masking of direct identifiers) is likely to be characterised as personal data. Therefore requirements of the EU General Data Protection Regulation ('GDPR' to come into effect on 25 May 2018) should be taken into account.
- Putting in place re-identification risk mitigatory measures will facilitate compliance with the GDPR.
- Data Pitch welcomes pseudonymised data provided that (1) re-identification becomes "reasonably" impossible from the pseudonymised datasets for sharing, and (2) any secondary processing of such datasets by a SME recipient would be compliant with the GDPR. Consequently:
 1. Technical and organisational measures should be put in place by the data provider to make sure the SME does not have access to the additional information required for recovering masked identifiers.
 2. The pseudonymised datasets for sharing with the SME should be stored on company-secure servers, or on the servers of the University of Southampton.
 3. The SME should only be permitted to process the pseudonymised datasets for sharing for a pre-specified analytics-driven purpose(s). Any indirect identifiers in the pseudonymised datasets for sharing should be removed or masked where these would not be strictly necessary for the SME to achieve the specified purpose(s). Similarly, possibilities for linking or inferring new information about data subjects from analysing the datasets – which could increase their risk of re-identification by the SME – should be muted as far as strictly necessary relative to the specified purpose(s).
 4. The specified purpose(s) should be compatible with the initial purpose(s) (for which the data was originally collected. Obtaining a clear description of the initial purpose and the legal basis justifying the initial collection and the challenge to be solved now will help in assessing this compatibility. To note reformulating the challenge so that the end result would be the production of statistics (aggregated data) - rather than decisions being taken about (that could affect) the data subjects - would ease compliance and a finding of compatibility of purposes.
 5. As data subjects have a right to object to secondary processing on data relating to them in certain circumstances, customers should be informed of what is being proposed where the scope of any initial consent obtained from them does not extend to the specified purpose(s). Even if consent was not obtained, but another legal basis for collection of personal data relied upon, it is still worth checking the breadth of any information provided to relevant individuals about the purpose why the data was collected and any comments made at that time about restrictions on the scope of that purpose.
 6. If the specified purpose(s) would involve data subjects being subject to 'profiling' analytics – and measures or decisions are subsequently planned to be taken vis-a-vis individual customers based on profiles that might be created by the SME - a data protection impact assessment (see the GDPR, Article 35) shall be undertaken before the measures or decisions are taken.
 7. To note, all SMEs will be required to comply with, and provided training on compliance with, data protection law under the GDPR. SMEs will also be required to destroy the data at the end of the Data Pitch project.