DESTINI

SMART DATA PROCESSING AND SYSTEMS OF DEEP INSIGHT

http://www.destini2020.eu

Deliverable D5.3

Report on content for training sessions 1

Document details:

Editor:	CUT
Contributors:	CUT, UNIROMA, JADS
Date:	28/09/2022
Version:	V6.0

Document history:

Version	Date	Contributor	Comments
1.0	17/06/2022	A.S. Andreou S. Mappouras A. Christoforou P. Christodoulou M. Pingos	Initial document, structure and content
2.0	20/07/2022	A.S. Andreou S. Mappouras	Additions & Corrections
3.0	01/08/2022	A.S. Andreou Partners	First review
4.0	05/09/2022	S. Mappouras	Additions & Corrections
5.0	15/09/2022	A.S. Andreou	Final review
6.0	28/09/2022	Partners	Final review & approval

Contents

1. Introduction

1.1. Purpose

This document presents a detailed description of the material used for performing training of researchers and hands-on experience sharing on methods, tools, and platforms for developing and assessing proof-of-concept prototypes and demos.

This deliverable is part of Work-Package 5 (WP5) that develops and executes a plan for joint education and training programmes between the members of the consortium. It also devises and applies a mobility programme to help attracting early-stage researchers within the consortium and beyond, describes training and mentoring activities offered to researchers and outlines incentives, like the announcement of PhD topics in high-demand and popular subjects, financial and local support for relocation.

The main focus of the mobility programme of the project in the context of WP5 was to attract early-stage researchers (ESRs), where they would travel to Cyprus for a period of time ranging from 1 to 4 weeks in order to enhance their research capabilities with the knowledge and expertise gained during the duration of the project. Despite the dissemination activities that took place, the consortium was unable to attract any ESRs during the initial period of the corresponding workpackage due to the COVID-19 virus outbreak, where travelling was forbidden while some countries enforced quarantine. The consortium held long teleconferencing discussions trying to find alternative ways to make its mobility program successful once any bans on travelling would be lifted, taking also into account that time was also a critical factor. In this context the partners indeed managed to pull this task through mainly by utilizing other events already planned and exchanging ESRs. More specifically:

The consortium organized a Satellite event to support and launch its the mobility program in the context of the 16th Symposium and Summer School On Service-Oriented Computing (SummerSoc

- https://www.summersoc.eu/) where it was facilitated in Crete, on July, 3-8, 2022. This school is a popular event among ESRs and it is organized each year with rich participation.

The aim of the DESTINI satellite event was to bring together ESRs for networking and collaboration in different scientific areas of interest which revolve around Data Science, with particular interest on Smart Data Processing and Systems of Deep Insight.

ESRs from the three partner universities and external ESRs participating at SummerSoc main event were invited to participate also in the DESTINI's satellite mobility event to present their research ideas and brainstorm with other ESRs with similar research interests. Subsequently, the consortium of DESTINI selected ESRs with the best research ideas that aligned with the project's Joint Research Areas to come to Cyprus for a short period of time, in order to continue their active involvement in the mobility program, and through this program to be able to collaborate with CUT's researcher on a specific topic, lay down prospective research ideas, work on them for producing research papers and thus enhance their research capabilities.

1.2. Definitions, Acronyms, and Abbreviations

ESRs: Early-stage researchers

MLOps: Machine Learning Operations

1.3. Overview

The rest of the document is structured as follows:

Section 2 presents briefly the dissemination material produced for the satellite event organized in Crete, consisting of the event invitation, posters and leaflets. Section 3 reports in details the training and brainstorming activities that were performed with speakers and presentations of different topics. Section 4 describes the selection of the best research ideas proposed by ESRs from Sapienza University of Rome, and Jheronymous Academy of Data Science for continuing with the mobility program by relocating in Cyprus. Finally, section 5 provides the concluding remarks.

2. Dissemination Material

2.1. Event Invitation

Call for Abstracts:

Presentation of Research Proposals

ESRs are invited to submit an extended abstract of 2-3 pages in which they will describe an **innovative idea for research or application** that falls within the scientific areas of the school. The abstract may include also some code, GUI screens or a graphical abstract to better demonstrate the idea. The submission may also involve proof-of-concept work or full applications with demos available.

The abstracts will be evaluated by the consortium and the most interesting ideas will be shortly presented, while brainstorming discussions will follow, which may span more than one day, to investigate whether an idea can be elaborated further to lead to a **potential research article for publication**. In this case the successful proposers will be invited to come to Cyprus for a period of 2-4 weeks, work their idea at the Cyprus University of Technology (*https://www.cut.ac.cy/?languageId=1*) with researchers from the consortium and write their papers. The invited ESRs will be **financially supported** by DESTINI covering the expenses of their air ticket and accommodation, plus a daily allowance for the period of their stay.

DESTINI Activity 1:
Collaborative Projects/Hackathon

ESRs are invited to a practice and competition activity, during which they will collaborate on specific (predefined) topics during the SummerSOC, with a kick-off on the first day (Monday) and with presentations on the last (Friday). Proposed solutions/ideas should/could be illustrated with some working prototypes. The ESRs may work in groups and will be guided and mentored by researchers participating in DESTINI. The topics include, but are not limited to, **smart data processing, data-driven decision support, data lakes and data meshes, Blockchain and Process Mining**. Successful solutions/ideas will be assessed by the consortium in terms of innovation and the potential to lead to a research article for publication. In such cases, the ESRs will be invited to come to Cyprus for a period of 2-4 weeks, work their idea at the Cyprus University of Technology (*https://www.cut.ac.cy/?languageId=1*) with researchers from the consortium and write their papers. The invited ESRs will be **financially supported** by DESTINI covering the expenses of their air ticket and accommodation, plus a daily allowance for the period of their stay.

DESTINI Activity 2:
Invitation for Continued Collaboration

ESRs that participate to SummerSOC and have successfully presented and defended their work in the main SummerSOC event are invited to participate to the **DESTINI Activity 2** for brainstorming and further exploration of their work. During this activity, DESTINI's consortium will suggest enhancements of their ideas with methods and techniques from the project's scientific areas of interest, such as, **smart data processing, data lakes and data meshes, Blockchain and Process Mining**. Successful development of joint ideas that could potentially lead to new **research articles** for publication will result in an invitation to the successful ESRs to come to Cyprus for a period of 2-4 weeks, work their idea further at the Cyprus University of Technology (https://www.cut.ac.cy/?languageId=1) with researchers from the consortium and write their papers. The invited ESRs will be **financially supported** by DESTINI covering the expenses of their air ticket and accommodation, plus a daily allowance for the period of their stay.

Participation and Submission

The interested ESRs may participate in one or more activities described above by notifying the organizers as follows:

Deadline for abstract submission	:	*June 15, 2022*
Notification of abstract acceptance	:	*June 22, 2022*
Declaration of interest in Activities 1, 2	:	*June 25, 2022*

Abstracts and declaration of interest must be submitted via email to one of the organizers:

Professor Andreas S. Andreou

Cyprus University of Technology
Department of Electrical Engineering, Computer Engineering and Informatics
Limassol, Cyprus
Phone: +357 25 00 2204,
Fax: +357 25 00 2635
E-mail: andreas.andreou@cut.ac.cy

Professor Willem-Jan van den Heuvel

Tilburg University
Jheronimus Academy of Data Science (JADS)
Tilburg, Netherlands
Phone: +31 13 466 2767
Fax: +31 13 466 3069
E-mail: W.J.A.M.v.d.Heuvel@jads.nl

Professor Massimo Mecella

Sapienza Università di Roma
Department of Computer, Control and Management Engineering
Rome, Italy
Phone: + 390677274028
Fax: + 390677274002
E-mail: mecella@dis.uniroma1.it

Consortium

Cyprus University of Technology

SAPIENZA UNIVERSITÀ DI ROMA

TILBURG UNIVERSITY

JADS
Jheronimus Academy of Data Science

destini2020.eu

facebook.com/destini2020eu

twitter.com/destini2020eu

2.2. Event Poster

2.3. Event Leaflet

DESTINI

Smart Data Processing and Systems of Deep Insight

Twinning | Horizon 2020

About Destini

DESTINI H2020 Twinning Project proposes a series of coordination and support actions for promoting research in the area of Smart Data. It brings together two internationally recognized scientific groups from the Netherlands (Tilburg University and Jheronymous Academy of Data Science - ERISS/JADS) and Italy (Sapienza Universita di Roma – UNIROMA1) that collaborate with Cyprus University of Technology (CUT) so as to strengthen CUT's research and scientific profile in the relevant area.

Aim

The aim of DESTINI is to facilitate transfer of scientific knowledge and expertise, as well as of best research practices from the leading institutions to CUT. The ultimate goal is that the research group of CUT increases its research capacity and prowess, by investigating a number of significant and hot topics in the field of Smart Data Processing and Systems of Deep Insight.

Cyprus University of Technology

SAPIENZA UNIVERSITÀ DI ROMA

TILBURG UNIVERSITY

JADS Jheronimus Academy of Data Science

Research Areas

1. Data Lakes

- Query sources and create SPARQL subgraphs using Visual Querying
- Characterize/ build blueprints of the data sources using Visual Querying
- Apply algorithms (classification) during the process of selecting sources based on their features and options
- Run the process continuously by renewing the subgraph
- Utilize this approach after data sources are ingested in a Data Lake
- Use blockchain in Data Lakes to address one of the most important challenges according to literature, that is, Security, Privacy and Data Governance
- Utilization of algorithms such as top k-means after data sources are ingested in the Data Lake

Cyprus University of Technology

SAPIENZA UNIVERSITÀ DI ROMA

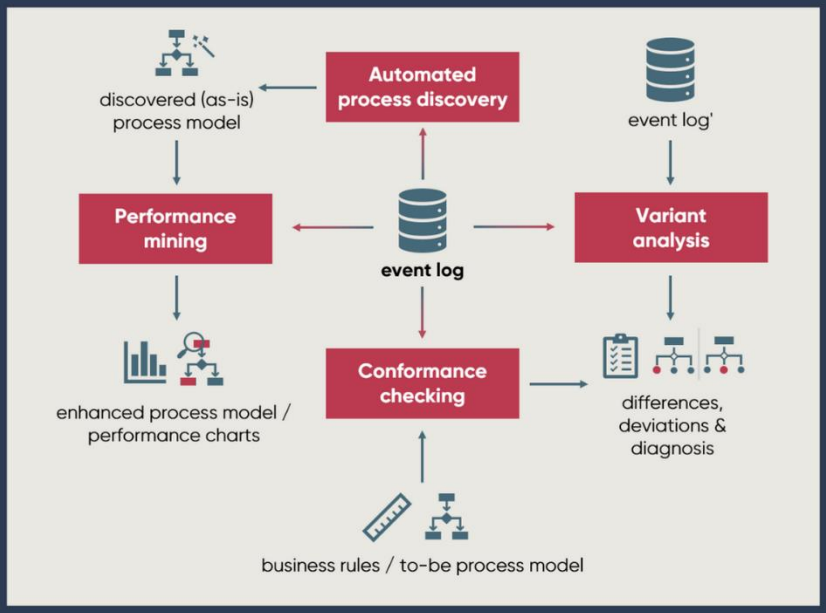
TILBURG UNIVERSITY

JADS
Jheronimus Academy of Data Science

Research Areas

2. Business Process Mining

- Perform process discovery and model creation
- Analysis of the business process based on event logs
- Conformance checking
- Pre-define goals for optimization
- Evolution - enhancement of a business process using machine learning
- Applications in smart manufacturing and/or the health sector

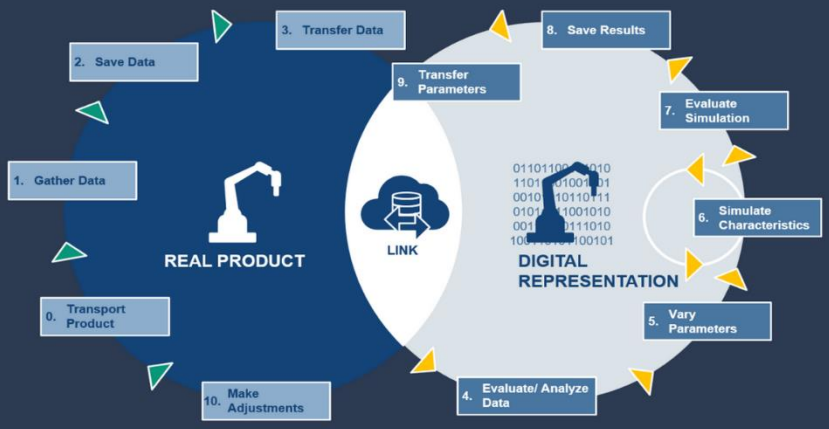


Cyprus University of Technology
SAPIENZA Università di Roma
TILBURG UNIVERSITY
JADS Jheronimus Academy of Data Science

Research Areas

3. Digital Twins

- Experimental digital twins for smart manufacturing
- Interactive visualization of physical entities
- Interaction between physical and virtual worlds
- Predictions based on real-world data from physical entities
- Graphical representation of real-world data on digital twins

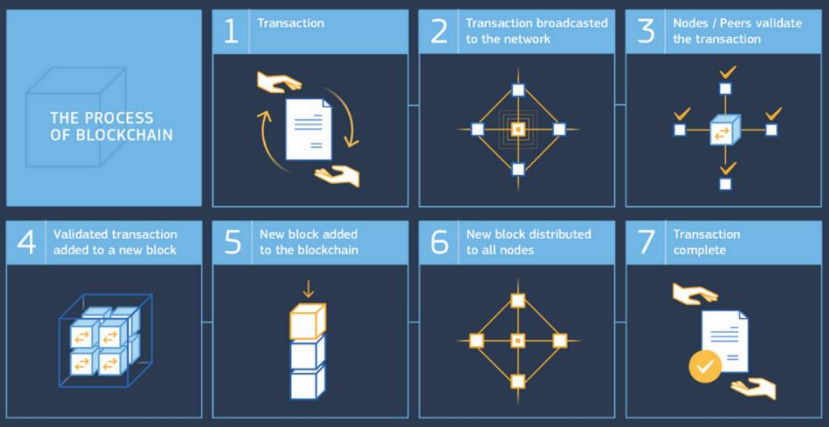


Cyprus University of Technology · Sapienza Università di Roma · Tilburg University · JADS Jheronimus Academy of Data Science

Research Areas

4. Blockchain

- Experimental Digital Twins in Blockchain
- Physical Entities as NFTs
- NFTs and Smart Contracts in a 3D environment
- Blockchain and BPM activities
- Blockchain and Data Lakes



Cyprus University of Technology

SAPIENZA UNIVERSITÀ DI ROMA

TILBURG UNIVERSITY

JADS Jheronimus Academy of Data Science

Consortium

Cyprus University of Technology

SAPIENZA UNIVERSITÀ DI ROMA

TILBURG UNIVERSITY

JADS
Jheronimus Academy of Data Science

destini2020.eu

facebook.com/destini2020eu

twitter.com/destini2020eu

2.4. Event Agenda

DESTINI SATELLITE EVENT - SUMMERSOC 2022

FINAL PROGRAM

Sunday, 03 July 2022, 10:00-14:00 – Preparation Day

10:00 – 14:00: Preparative activities for DESTINI's Satellite Event

Monday, 04 July 2022, 09:00-17:30 – Orientation Day

09:00 – 10:00: Presentation of the DESTINI project / Orientation, A.S. Andreou (CUT)

10:00 – 16:30: Talks (jointly with SummerSOC), brainstorming sessions and mentoring activities

16:30 – 17:30: Wrap-up of Orientation Day

Tuesday, 5 July 2022 – Thursday 7 July 2022, 09:00-17:30 – Working days #1 -- #3

09:30 – 10:00: Launch of the mentoring activities

10:00 – 16:30: Talks (jointly with SummerSOC), brainstorming sessions and mentoring activities

16:30 – 17:30: Wrap-up of Working Day #1 -- #3

Friday 8 July 2022 – Working day #4

16:30 – 18.30: Presentation and selection of best ideas for the mobility program

3. Training and Brainstorming Activities

3.1. DAY 1 - 04/07/2022

The first day of the event included one introductory presentation about the DESTINI project and the mobility program. Then, brainstorming sessions followed, with participating ESRs exchanging ideas on their research and its alignment with DESTINI's areas of interest.

3.1.1. DESTINI Project Overview, Mobility Program - A.S. Andreou

Presenter:	Prof. Andreas S. Andreou	**Department / Division:**	Cyprus University of Technology
Subject:	DESTINI Project Overview, Mobility Program		
Date:	04/07/2022		

1. Subject short description
Professor Andreou presented an overview of the project's outcomes thus far, explaining the project's JRA's and project's objectives. In addition, the professor explained the idea of the mobility program, and as the coordinator of the event gave directions and guidelines for the remaining time of the event.

Attendees of Day 1 - 04/07/2022		
Name	**Email**	**Department/Division**
Andreas Andreou	andreas.andreou@cut.ac.cy	CUT
Panayiotis Christodoulou	paa.christodoulou@edu.cut.ac.cy	CUT
Stelios Mappouras	sm.mappouras@edu.cut.ac.cy	CUT
Andreas Christoforou	andreas.christoforou@cut.ac.cy	CUT
Indika Kumara	i.p.k.weerasinghadewage@tue.nl	JADS
Stefan Driessen	s.w.driessen@jads.nl	JADS
Stefano Dalla Palma	s.dallapalma@uvt.nl	JADS
Mirella Sangiovanni	m.sangiovanni@tue.nl	JADS
Daniel De Pascale	d.de.pascale@tue.nl	JADS
Nemania Borovits	n.borovits@jads.nl	JADS
Damian A. Tamburri	d.a.tamburri@tue.nl	JADS
Massimo Mecella	mecella@diag.uniroma1.it	SAPIENZA
Jerin George Mathew	mathew@diag.uniroma1.it	SAPIENZA
Alberto Morvillo	morvillo@diag.uniroma1.it	SAPIENZA
Silvestro Veneruso	veneruso@diag.uniroma1.it	SAPIENZA

Francesco Leotta	leotta@diag.uniroma1.it	SAPIENZA
Flavia Monti	monti@diag.uniroma1.it	SAPIENZA
Willem-Jan van den Heuvel	w.j.a.m.v.d.heuvel@jads.nl	JADS

3.2. DAY 2 - 05/07/2022

Day 2 of the event included 3 presentations, one from an ESR from SAPIENZA, and two from Senior researchers of CUT, which were followed up with mentoring activities and brainstorming.

3.2.1. Intelligence in Industry 4.0 - Flavia Monti

Presenter:	Flavia Monti	**Department / Division:**	SAPIENZA
Subject:	Intelligence in Industry 4.0		
Date:	05/07/2022		

1. Subject short description
Industry 4.0, is a term referring to the fourth industrial revolution of manufacturing processing and practices. Industry 4.0 aims at complete digitization of industrial automated production and stimulates the usage of many innovative technologies such as Artificial Intelligence (AI), Data Mining, Big Data Analytics (BDA) and Cloud Computing for the analysis of such data to increase productivity and guarantee production quality. With the advent of Industry 4.0, diagnostic and prognostic models have become more common. In particular, such models aim at detecting unusual and anomalous behavior in the production, extract information from failures and anomalies, and estimate and predict the future (e.g., Remaining Useful

Life (RUL)). Closely related to these models is the implementation of Predictive Maintenance (PdM) strategies, which have become an indispensable trend to obtain performance degradation information.

Digital Twins (DTs) are considered one of the enabling key factors of Industry 4.0. DTs are nowadays intended as digital representations of physical entities during their entire lifecycle. DTs are commonly used for different purposes, for instance the control problem, i.e., the problem of selecting the action to do next. Particularly, DTs and planning techniques can be joined to set up reliable production plans.

It is clear how the transition to Industry 4.0 requires absolute knowledge of many technologies. Because of the multitude of solutions and techniques, it is not easy to plan the roadmap to shift towards Industry 4.0 and it is not even explicit for a company to understand its readiness as an Industry 4.0 player. The presence of a model to assess the maturity and readiness of a company as an Industry 4.0 actor allows them to evaluate their strengths and weaknesses.

3.2.2. Models for microservices adoption - Dr. Andreas Christoforou

Presenter:	Dr. Andreas Christoforou	**Department / Division:**	Cyprus University of Technology
Subject:	Models for microservices adoption		
Date:	05/07/2022		

1. Subject short description

Cloud computing has been established as a new development environment that simplifies the procedure of building and hosting applications. This has inevitably presented new challenges in the field of software engineering. The persistent incorporation of numerous new technologies makes Cloud infrastructure management highly complex, with multi-conflicting factors affecting it. Computational intelligence and machine learning techniques appear to succeed when dealing with

complex and multifaceted problems. Our latest research developments revolve around utilizing computational intelligence techniques and approaches that are modified and extended to meet the needs of specific challenges. A particular reference is made to a proposed novel integrated analysis framework based on Multi-Layer Fuzzy Cognitive Maps models and a series of actions to gather useful static and dynamic information. The aforementioned framework utilizes techniques based on Artificial neural networks, Fuzzy logic, and Evolutionary computation. The strength and applicability of the proposed model were demonstrated through the support of the decision to adopt microservices architecture in a more sophisticated form with enhanced explainability and interpretability features. Although the decision to adopt the microservices architecture is extremely tough, the model captured the dynamics of the problem under study, provided a better understanding of the environment under study, and explained the produced results.

3.2.3. Blockchain in Data Lakes and other applications - Dr. Panayiotis Christodoulou

Presenter:	Dr. Panayiotis Christodoulou	**Department / Division:**	Cyprus University of Technology
Subject:	Blockchain in Data Lakes and other applications		
Date:	05/07/2022		

1. Subject short description

Nowadays, as stated by the EU, Green bonds play an increasingly supreme role in financing and supporting climate-related or environmental projects. The main questions that arise now is on how we can ensure that EU green bonds are invested for green growth purposes and not for greenwashing and on how we can monitor the transaction process. This presentation initially, outlines how the scientific literature deals with greenwashing and then it presents EUGBS-Chain, a blockchain-based framework for effective monitoring of EU Green Bonds that aims to increase transparency and accountability eliminating at the same time the risk of greenwashing. The proposed framework has important implications for all involved stakeholders such as issuers, investors and policy makers as it reduces monitoring costs, minimizes greenwashing and improves effectiveness, transparency, comparability and credibility of the market.

2. Attendees of Day 2 - 05/07/2022		
Name	**Email**	**Department/Division**
Andreas Andreou	andreas.andreou@cut.ac.cy	CUT
Panayiotis Christodoulou	paa.christodoulou@edu.cut.ac.cy	CUT
Stelios Mappouras	sm.mappouras@edu.cut.ac.cy	CUT
Andreas Christoforou	andreas.christoforou@cut.ac.cy	CUT
Indika Kumara	i.p.k.weerasinghadewage@tue.nl	JADS
Stefano Dalla Palma	s.dallapalma@uvt.nl	JADS
Mirella Sangiovanni	m.sangiovanni@tue.nl	JADS
Daniel De Pascale	d.de.pascale@tue.nl	JADS
Nemania Borovits	n.borovits@jads.nl	JADS
Damian A. Tamburri	d.a.tamburri@tue.nl	JADS
Massimo Mecella	mecella@diag.uniroma1.it	SAPIENZA
Jerin George Mathew	mathew@diag.uniroma1.it	SAPIENZA
Alberto Morvillo	morvillo@diag.uniroma1.it	SAPIENZA

Silvestro Veneruso	veneruso@diag.uniroma1.it	SAPIENZA
Francesco Leotta	leotta@diag.uniroma1.it	SAPIENZA
Flavia Monti	monti@diag.uniroma1.it	SAPIENZA
Willem-Jan van den Heuvel	w.j.a.m.v.d.heuvel@jads.nl	JADS

3.3. DAY 3 - 06/07/2022

Day 3 of the event included 6 presentations, one from an ESR of CUT, and five ESRs from JADS. Later on, mentoring activities and brainstorming were performed.

3.3.1. Blockchain and NFTs: Challenges and solutions - Stelios Mappouras

Presenter:	Stelios Mappouras	**Department / Division:**	Cyprus University of Technology
Subject:	Blockchain and NFTs: Challenges and solutions		
Date:	06/07/2022		

1. Subject short description
Stelios Mappouras (MSc candidate and researcher on DESTINI) presented his research idea for his MSc thesis, which is about combining Blockchain and NFT technologies with Digital Twins. The first part is about an NFT marketplace with NFTs as physical items, and the second part about developing an NFT infrastructure which will host several NFT applications, including the combination of NFTs with digital twins, etc.

3.3.2. Data Product marketplaces - Stefan Driessen

Presenter:	Stefan Driessen	**Department / Division:**	JADS
Subject:	Data Product marketplaces		
Date:	06/07/2022		

1. Subject short description
The research activities of Stefan Driessen revolve around alternatives to monolithic data platforms such as data lakes and data warehouses. When dealing with heterogeneous data sources and use cases across different (business) domains, the centralized architectures prove to be a bottleneck that hinders the scaling of effective data exchange. Alternative platforms, such as data mesh and data markets, incorporate a left shift in responsibility of data provisioning. In their architectures, domain teams provide data products that they optimize for consumption. Currently, he is investigating the implications of the `decentralisation' of data exchanging required by these platforms.

3.3.3. MLOps, testing - Stefano Dalla Palma

Presenter:	Stefano Dalla Palma	**Department / Division:**	JADS
Subject:	MLOps, testing		
Date:	06/07/2022		

1. Subject short description

Infrastructure as Code (IaC) is the DevOps strategy enabling management and provisioning of infrastructure through the definition of machine-readable files, rather than physical hardware configurations. As any other source code artifact, these files may, however, contain defects that can preclude their correct functioning.

Little attempts have been conducted to define learning models that predict the defectiveness of IaC files.

In the context of his PhD, one of the main concerns regard the understanding of the characteristic of defective IaC scripts and defining models to detect them, as well as a tool suite for IaC defect prediction in practice. The PhD was funded by the European Union's Horizon research and innovation project called RADON (https://radon-h2020.eu).

The project aims at pursuing a broader adoption of serverless computing technologies within the European software industry.

One of RADON's key pillars is quality assurance tools for Infrastructure as Code.

As part of the RADON framework, the Defect Prediction tool (https://github.com/radon-h2020/radon-defuse) developed during the PhD focuses on IaC correctness, with a focus on the configuration management language Ansible and the OASIS standard for cloud applications orchestration, called TOSCA.

The tool consists of several components to mine open-source repositories to identify failure data and to extract quality metrics to guide the empirical training and enrichment of the models for defect prediction as well as predict code smells and errors in IaC files.

Although instantiated in the context of infrastructure as code, specifically for Ansible and TOSCA, the defect predictor was envisioned and designed to be language-agnostic.

3.3.4. Crime Analytics & Deep Web - Daniel De Pascale

Presenter:	Daniel De Pascale	**Department / Division:**	JADS
Subject:	Crime Analytics & Deep Web		
Date:	06/07/2022		

1. Subject short description

Considering the massive increase in the number of crimes in the last decade, as well as the outlook toward smarter cities and more sustainable urban living, the emerging cyber-physical space (CPS) obtained by the interaction of such physical spaces with the cyber elements around them plays a key role in the protection of urban social living. This project aims to outline a vision where CPS protection is center-stage and where CPS models as well as so-called hybrid analytics work jointly to help the Law Enforcement Agents (LEAs), e.g., in event monitoring and early detection of criticalities. As a part of the project, we implement a case study in the scope of VISOR, a Dutch government project aimed at improving CPS protection using hybrid analytics. We conduct a Field experiment in the Paaspop social event and festival grounds to test and select the most appropriate device configuration. Then, we create and validate the KGen middleware, a prototype tool to anonymize structured big data using genetic algorithms, and SENSEI, a frame work for dark web market Place analytics. We conclude that hybrid analytics offer a considerable ground for more sustainable CPS.

3.3.5. Privacy Preservation - Nemania Borovits

Presenter:	Nemania Borovits	**Department / Division:**	JADS
Subject:	Privacy Preservation		
Date:	06/07/2022		

1. Subject short description

In the past decade we have witnessed a major in advancement of AI mainly driven by Big Data (volume, variety, velocity and veracity), powerful graphics processing unit (GPU) and breakthroughs in machine learning, deep learning, reinforcement learning, and other techniques. Driven by the introduction of the corresponding legislation introduced around the world, it is a new challenge to discover AI knowledge from big data while not compromising data security and privacy. Based on the Privacy by Design framework various Privacy Enhancing Technologies (PETs) have been introduced. We presented a study which focuses on Federated Learning (FL). FL is a framework to train a Machine/Deep Learning model in a distributed privacy-preserving manner. The main idea is for the data to remain at their source where a local model is trained with the purpose of centrally aggregating just the parameters to build a common global model. This new scheme allowed ML and AI practitioners to circumvent many of the legal restrictions and created new opportunities in distributed applications where centrally aggregating data would be very inefficient. In our study we performed a short Systematic Literature Review (SLR) in order to identify design patterns for the FL main architecture in order to ensure better privacy-privacy preservation while not compromising data utility. In addition, we implemented one of the identified design patterns, namely client clustering. We conducted an experiment using 3 datasets implementing a Named Entity Recognition (NER) modelling task within the context of FL. Our results suggest that incorporating FL results in better model performance for the corresponding task. Furthermore, we demonstrated that the results of the modelling task (NER) yield further improvement after adding the developed design pattern i.e. client clustering.

3.3.6. Sustainable cities and Smart Healthcare Intelligence - Mirella Sangiovanni

Presenter:	Mirella Sangiovanni	**Department / Division:**	JADS
Subject:	Sustainable cities and Smart Healthcare Intelligence		
Date:	06/07/2022		

1. Subject short description

As Ph.D. at JADS and participant to DESTINI project, Mrs. Mirella gives her contributions to explore some relevant and critical aspects like Smart Healthcare Intelligence, Crime Security & Safety, Nature & AgriFood, and Sustainability. These domains align with the actual objectives and issues to improve the quality of life and reduce manual and time-consuming approaches.

In particular, in the context of Forensics investigations, Mrs. Mirella explores new approaches and implementations to support traditional manual methods in identifying genetic profiles - e.g., convolutional neural network to improve electropherogram classification performance.The difficulty in interpretation depends on abnormalities occurring in the laboratory steps. It is translated into problems and difficulties in the interpretation of the results and possible wrong identification of matching between DNA profiles and consequently mistakes about the people involved in the crime scene.

2. Attendees of Day 3 - 06/07/2022		
Name	**Email**	**Department/Division**
Andreas Andreou	andreas.andreou@cut.ac.cy	CUT
Stelios Mappouras	sm.mappouras@edu.cut.ac.cy	CUT
Andreas Christoforou	andreas.christoforou@cut.ac.cy	CUT
Indika Kumara	i.p.k.weerasinghadewage@tue.nl	JADS
Stefan Driessen	s.w.driessen@jads.nl	JADS
Stefano Dalla Palma	s.dallapalma@uvt.nl	JADS
Mirella Sangiovanni	m.sangiovanni@tue.nl	JADS
Daniel De Pascale	d.de.pascale@tue.nl	JADS
Nemania Borovits	n.borovits@jads.nl	JADS
Damian A. Tamburri	d.a.tamburri@tue.nl	JADS
Massimo Mecella	mecella@diag.uniroma1.it	SAPIENZA
Jerin George Mathew	mathew@diag.uniroma1.it	SAPIENZA
Alberto Morvillo	morvillo@diag.uniroma1.it	SAPIENZA

Silvestro Veneruso	veneruso@diag.uniroma1.it	SAPIENZA
Francesco Leotta	leotta@diag.uniroma1.it	SAPIENZA
Flavia Monti	monti@diag.uniroma1.it	SAPIENZA
Willem-Jan van den Heuvel	w.j.a.m.v.d.heuvel@jads.nl	JADS
Michalis Pingos	michalis.pingos@cut.ac.cy	CUT
Spyros Loizou	spyros.loizou@cut.ac.cy	CUT

3.4. DAY 4 - 07/07/2022

Day 4 of the event included 5 presentations, two from ESRs of CUT, and three ESRs from SAPIENZA. Later on, mentoring activities and brainstorming were performed.

3.4.1. Applying Process Mining to Human Daily Activities - Silvestro Veneruso

Presenter:	Silvestro Veneruso	**Department / Division:**	SAPIENZA
Subject:	Applying Process Mining to Human Daily Activities		
Date:	07/07/2022		

1. Subject short description

The context of the presentation was smart spaces, specifically, environments enriched with sensors and actuators, which aim to provide automatic/semi-automatic services to its inhabitants. A smart space produces data in the form of a sensor log, which consists in a series of raw measurements from the sensors within the smart environment itself. Researchers need open datasets to validate their approaches, thus ensuring reproducibility. Unfortunately, most of the datasets that we can find in the literature suffer from a set of limitations, e.g., they consider only specific categories of sensors or are designed to highlight the features of approaches proposed by the authors. As first contribution, the ESR have created a tool to produce synthetic datasets for the smart space community. It is based on human activity models and it produces logs in the XES format, i.e., a well-known IEEE standard based on XML.

During his PhD, he focused on developing unsupervised learning techniques to mine human habits. The term habit denotes a human routine, e.g., “what the user usually does in the morning after breakfast?”. The goal is to provide autonomous mechanisms to (i) segment logs (in our case we start from sensor logs), with the aim of reducing the human effort and avoiding annoying and sometimes

imprecise training sessions, (ii) discover models of human habits through Process Mining, and (iii) to use this model for anomaly detection and recommendations.

3.4.2. Geo-based visual exploration of digital document collections - Alberto Morvillo

Presenter:	Alberto Morvillo	**Department / Division:**	SAPIENZA
Subject:	Geo-based visual exploration of digital document collections		
Date:	07/07/2022		

1. Subject short description
Geo-based visual exploration of digital document collections Content exploration for digital humanities poses many challenges deeply connected to the user behaviours - Data sources, which may not be mean for a digital platform - Cross-references and implied information, where contents may be referred by each other and/or implies implicit concepts - Data retrieval and visualization, where researchers may require to deal with digital content in a similar way they does with non-digital ones - in particular in the field of Archaeology, a reference to a geographic location is also involved SCIBA is a platform currently in development which allows researchers in the field of archaeology to explore knowledge bases with cartographic references and semantic analysis. It uses a triplestore combined with a map service and GeoJSON documents to provide a

geo-based visual exploration of a digital document collection.

3.4.3. Statistical and Machine Learning Approaches to Record Linkage - Jerin George Mathew

Presenter:	Jerin George Mathew	**Department / Division:**	SAPIENZA
Subject:	Statistical and Machine Learning Approaches to Record Linkage		
Date:	07/07/2022		

1. Subject short description

Record Linkage (RL), also known as Entity Resolution (ER), aims at identifying pairs of records from different data sources that refer to the same real-world entity. RL is arguably the most important step in the data integration pipeline, and several approaches have been proposed in the literature, based either on statistical methods or more recent Machine Learning (ML) models.

These two approaches are at the two ends of a broad range of RL methods. Statistical approaches are mostly unsupervised methods and are transparent, that is, the inner logic behind these techniques is reasonably accessible and understandable to the end user.

On the other hand, ML techniques are thought to be more accurate than statistical approaches. Still, they are usually opaque, i.e., it may be challenging to understand the rationale behind their predictions.

3.4.4. Data Lakes and semantic enrichment mechanisms - Michalis Pingos

Presenter:	Michalis Pingos	**Department / Division:**	CUT
Subject:	Data Lakes and semantic enrichment mechanisms		
Date:	07/07/2022		

1. Subject short description

One of the greatest challenges in Smart Big Data Processing nowadays revolves around handling multiple heterogeneous data sources that produce massive amounts of structured, semi-structured and unstructured data through Data Lakes. The latter requires a disciplined approach to collect, store and retrieve/ analyse data to enable efficient predictive and prescriptive modelling, as well as the development of other advanced analytics applications on top of it. The present paper addresses this highly complex problem and proposes a novel standardization framework that combines mainly the 5Vs Big Data characteristics, blueprint ontologies and Data Lakes with ponds architecture, to offer a metadata semantic enrichment mechanism that enables fast storing to and efficient retrieval from a Data Lake. The proposed mechanism is compared qualitatively against existing metadata systems using a set of functional characteristics or properties, with the results indicating that it is indeed a promising approach.

3.4.5. Digital Twins for visual querying and process mining - Spyros Loizou

Presenter:	Spyros Loizou	**Department / Division:**	CUT
Subject:	Digital Twins for visual querying and process mining		
Date:	07/07/2022		

1. Subject short description
This presentation introduces a framework for business technology that combines the notion of Digital Twins with Process Mining aiming at delivering a simple and efficient way to retrieve customized data and process it with the use of graphical techniques, providing interactive visualization of process mining steps. More specifically, the proposed framework provides the ability to define different data sources and link these sources with a visual query generator which constructs, executes and depicts graphically the results of custom queries. The framework includes also sophisticated Artificial Intelligence / Machine Learning algorithms for data analysis, filtering and prediction. The framework is demonstrated through an interactive dashboard, which was implemented in Python to support a fully operational and visual process mining environment that facilitates decision making without the need of programming or data management skills.

2. Attendees of Day 4 - 07/07/2022		
Name	**Email**	**Department/Division**
Andreas Andreou	andreas.andreou@cut.ac.cy	CUT
Stelios Mappouras	sm.mappouras@edu.cut.ac.cy	CUT
Indika Kumara	i.p.k.weerasinghadewage@tue.nl	JADS
Stefan Driessen	s.w.driessen@jads.nl	JADS
Stefano Dalla Palma	s.dallapalma@uvt.nl	JADS
Daniel De Pascale	d.de.pascale@tue.nl	JADS
Nemania Borovits	n.borovits@jads.nl	JADS
Massimo Mecella	mecella@diag.uniroma1.it	SAPIENZA
Jerin George Mathew	mathew@diag.uniroma1.it	SAPIENZA
Alberto Morvillo	morvillo@diag.uniroma1.it	SAPIENZA
Silvestro Veneruso	veneruso@diag.uniroma1.it	SAPIENZA
Francesco Leotta	leotta@diag.uniroma1.it	SAPIENZA
Flavia Monti	monti@diag.uniroma1.it	SAPIENZA

Willem-Jan van den Heuvel	w.j.a.m.v.d.heuvel@jads.nl	JADS
Geert Monsieur	g.monsieur@jads.nl	JADS
Michalis Pingos	michalis.pingos@cut.ac.cy	CUT
Spyros Loizou	spyros.loizou@cut.ac.cy	CUT

3.5. DAY 5 - 08/07/2022

Day 5 of the event included 3 presentations from ESRs of SAPIENZA. Later on, mentoring activities and brainstorming sessions followed. Finally, a selection of the best research ideas was performed with ESRs being invited to come to Cyprus to continue their engagement with the mobility program and continue their research collaboration with CUT's researchers.

3.5.1. Big Data Pipelines Discovery from Dark Data: Challenges and Some Solutions - Andrea Marrella

Presenter:	Dr. Andrea Marrella	**Department / Division:**	SAPIENZA
Subject:	Big Data Pipelines Discovery from Dark Data: Challenges and Some Solutions		
Date:	08/07/2022		

1. Subject short description
Big Data pipelines are essential for leveraging Dark Data, i.e., data collected but not used and turned into value. However, tapping their potential requires going beyond existing approaches and frameworks for Big Data processing. The Computing Continuum enables new opportunities for managing Big Data pipelines concerning efficient management of heterogeneous and untrustworthy resources. The research of Andrea Marrella discusses the Big Data pipelines lifecycle on the Computing Continuum, its associated challenges and outlines a future research agenda in this area.

3.5.2. Event Log Extraction and Generation in the context of Big Data Pipeline - Dario Benvenuti

Presenter:	Dario Benvenuti	**Department / Division:**	SAPIENZA
Subject:	Event Log Extraction and Generation in the context of Big Data Pipeline		
Date:	08/07/2022		

1. Subject short description

Dario Benvenuti in the context of his presentation he introduced Process Mining, which is a discipline that sits between data mining and business process management. The starting point of process mining is an event log, which is analyzed to extract useful insights and recurrent patterns about how processes are executed within organizations. However, often its concrete application is hampered by the considerable preparation effort that needs to be conducted by human experts to collect the required data for building a suitable event log. Instead, event logs need to be extracted from different and heterogeneous data sources, often using customized extraction scripts whose implementation requires both technical and domain expertise. To tackle this issue Dario proposed an interactive and general-purpose approach to support organizations in generating simulated event logs that can be employed to discover the structure of the data pipelines (composite workflows for processing data that is enacted as part of process execution) executed within a business process. Furthermore, Dario is investigating the possibility of using queries directly on databases to easily perform Process Discovery.

Slides

3.5.3. KPI-driven process mining in E-government - Francesca De Luzi

Presenter:	Francesca De Luzi	**Department / Division:**	SAPIENZA
Subject:	Digital Twins for visual querying and process mining		
Date:	08/07/2022		

1. Subject short description

The control of the times, volumes and resources used in the treatment of judicial processes is a fundamental activity to obtain a better use of the same. In this regard, it is necessary to define Key Performance Indicators (KPIs) of the times and resources used as well as their visualization and accessibility to users.

The methodology investigated for the monitoring of judicial processes provides for the identification of appropriate KPIs through:

· State-of-the-art analysis on KPI extraction methodologies, with a particular focus on process KPIs, together with the study of Natural Language Processing (NLP) and Process Mining techniques to extract information from judgments and processes.

· Analysis of some judgments and identification of a use case in the context of civil trial.

· Understanding of macro-phases, activities, and events over time.

· Use case modeling through the production of a BPMN scheme.

· Extraction of temporal information in the legal field.

· Mapping between the activities of the trial and the temporal information obtained from the judgments to identify new indicators of process performance.

The case study concerns the contentious divorce process. In detail, the BPMN scheme shows the phases of the process from which useful temporal information was extracted. It has thus become clear that the case is delayed in the phase of issuing the decree, since this time interval (between the filing of the appeal and the issuance of the decree by the president of the court) must be set at five days. To date, the research activity has focused on the extraction of temporal information to identify new KPIs. However, further factors for their identification will be investigated in the future.

2. Attendees of Day 5 - 08/07/2022		
Name	**Email**	**Department/Division**
Stelios Mappouras	sm.mappouras@edu.cut.ac.cy	CUT
Savvina Rousou	sc.rousou@edu.cut.ac.cy	CUT
Stefan Driessen	s.w.driessen@jads.nl	JADS
Massimo Mecella	mecella@diag.uniroma1.it	SAPIENZA
Jerin George Mathew	mathew@diag.uniroma1.it	SAPIENZA
Alberto Morvillo	morvillo@diag.uniroma1.it	SAPIENZA
Silvestro Veneruso	veneruso@diag.uniroma1.it	SAPIENZA
Francesco Leotta	leotta@diag.uniroma1.it	SAPIENZA
Flavia Monti	monti@diag.uniroma1.it	SAPIENZA

Michalis Pingos	michalis.pingos@cut.ac.cy	CUT
Spyros Loizou	spyros.loizou@cut.ac.cy	CUT
Andrea Marrella	marrella@diag.uniroma1.it	SAPIENZA
Dario Benvenutti	d.benvenutti@diag.uniroma1.it	SAPIENZA
Francesca De Luzi	deluzi@diag.uniroma1.it	SAPIENZA

4. Selection of the best research ideas for the mobility program

During the last day of the event, the best research ideas were selected and ESRs were invited to come to Cyprus in the context of the mobility program and enhance their research capabilities in collaboration with researchers from CUT, with the aim being to produce research papers and/or prepare the ground for future proposals to attract further funding.

The ERSs that were selected along with their broad topic of interest were as follows:

4.1. Flavia Monti - Intelligence in Industry 4.0

Ideas expressed during the presentation:

- Decision support system using FCM:
- Study the dynamics contributing into taking decisions according to the readiness level of industry 4.0
- Predictive Maintenance for Maritime using Digital Twins

Potential ideas when visits Cyprus:

- Start developing an FCM in collaboration with Andreas Christoforou (post-doc researcher of CUT team with expertise on FCMs)
- Consider digital twins in maritime
- Elaborate with Spyros Loizou on Digital Twins (a PhD student from CUT also working on Digital Twins)
- Collaborate with JADS on a project they have "Shop Floor Monitoring System"

4.2. Silvestro Veneruso

Ideas expressed during the presentation:

- Combination of your idea with the "Crime Analytics" project from JADS (regarding public spaces)
- Write a proposal for funding based on this idea for civil protection, which will be a joint proposal from the 3 universities (SAPIENZA, JADS, CUT)

Potential ideas when visits Cyprus:

- Adapt the already developed learning techniques to the context of a smart factory
- Combine our ideas and techniques with the "Crime Analytics" project from JADS (regarding public spaces)
- write a proposal for fundings based on this idea for civil protection, which will be a joint proposal from the 3 universities (SAP, JADS, CUT).

4.3. Dario Benvenuti

Ideas expressed during the presentation:

- Assessment of the accuracy of the logs
- Secure the logs using blockchain technologies

Potential ideas when visits Cyprus:

- Association of the Event Log extraction with Data Lakes
- Application of Smart Data processing techniques such as FCM to event log extraction techniques

4.4. Jerin George Mathew

Ideas expressed during the presentation:

- Application of this statistical techniques for comparison on identifying similar/identical event logs for business process mining
- Use of the methodology for real-world application to identify similar situations for predictive maintenance

Potential ideas when visits Cyprus:

- Extension on the statistical methods for comparison with FCM to assist with decision support
- Utilization of the methodology to remove identical data from data lakes

4.5. Francesca De Luzi

Ideas expressed during the presentation:

- Apply the modelling techniques in more graphs such as Petri Net in order to achieve broader results
- Expansion of the data ingestion procedure from a broader data source such as data lakes so as to have more data, to extract more KPIs

Potential ideas when visits Cyprus:

- Utilization of FCMs to enhance the KPI identification procedure
- Visualization of the models using Digital Twins

4.6. Alberto Morvillo

Ideas expressed during the presentation:

- Use of the visual exploration on Digital Twins
- Utilize his geo-based visual exploration on Digital Cultural Heritage

Potential ideas when visits Cyprus:

- Collaboration with the Digital Cultural Heritage lab on the creation of knowledge graphs utilizing data lakes as source of the knowledge graph
- Utilization of the geo-based visual exploration on metaverse and NFTs

4.7. Mirella Sangiovanni

Ideas expressed during the presentation:

- Adoption of FCMs in her research
- Research about Bacteria in the milk

Potential ideas when she visits Cyprus:

- Adoption of FCMs in her research
- Research about Bacteria in the milk

4.8. Stefan Driessen

Ideas expressed during the presentation:

- Utilization of decentralized ledgers (blockchain) to data meshes for the security of the data
- Conduction of European/National proposal for future data platforms

Potential ideas when visits Cyprus:

- Extension of the work that has been done in Cyprus on data lakes with data meshes
- Collaboration on research for blockchain

5. Conclusions

This deliverable is part of Work-Package 5 (WP5) that develops and executes a plan for joint education and training programmes between the members of the consortium. Its main outcome is to devise and apply a mobility programme to help attracting early-stage researchers within the consortium and beyond, and describe training and mentoring activities offered to researchers and outline incentives, like the announcement of PhD topics in high-demand and popular subjects, scholarships and local support for relocation.

The mobility program was launched through a satellite in the context of the 16th Symposium and Summer School On Service-Oriented Computing (SummerSoc - https://www.summersoc.eu/), a popular event among ESRs and it is organized each year with rich participation, which was held in Crete, Greece on July, 3-8, 2022.

The aim of the DESTINI satellite event was to bring together ESRs for networking and collaboration in different scientific areas of interest which revolve around Data Science, with particular interest on Smart Data Processing and Systems of Deep Insight.

ESRs from the three partner universities and external ESRs participating at SummerSoc main event were invited to participate also in the DESTINI's satellite mobility event to present their research ideas and brainstorm with other ESRs with similar research interests. Subsequently, the

The event was concluded successfully, which is indicated by the rich participation and interest on behalf of the ESRs.