Ref. Ares(2020)2333125 - 01/05/2020

DESTINI

SMART DATA PROCESSING AND SYSTEMS OF DEEP INSIGHT

http://www.destini2020.eu

Deliverable D2.3

Research & Innovation Agenda

Document details:

Editor:	Mike Papazoglou
Contributors:	Andreas Andreou, Panayiotis Christodoulou
Date:	April 30, 2020
Version:	6.0

Document history:

Version	Date	Contributor	Comments
v.1.0	10/4/2020	Mike Papazoglou	First Draft
v.2.0	13/4/2020	Andreas Andreou	Comments & Corrections
v.3.0	17/4/2020	Mike Papazoglou	Completed version
v.4.0	23/4/2020	Andreas Andreou Panayiotis Christodoulou	Corrections
v.5.0	27/4/2020	Mike Papazoglou	Final Corrections
v.6.0	30/4/2020	Partners	Approved final version after minor revision

Contents

1 Introduction

The DESTINI project is ambitious and realistic about what is needed to stimulate the uptake of Smart Data Processing and Systems of Deep Insight and about how and where it can contribute to the widening institution. Its primary strength comes from aligning between the widening research institution and leading partners into a synergistic research ecosystem, rooted in the integration of research areas that encompass AI, smart data, software services and IoT, and in exploring synergies, build on the huge potential of networking for excellence through knowledge transfer and, exchange of best research practices in the important field of Smart Data Processing and Systems of Deep Insight research. This partnership provides a unique opportunity to create greater value for European research and society through the activities of the network.

The main thrust of DESTINI's research is to deliver the research instruments and technology of Smart Data Processing and Systems of Deep Insight and to provide efficient and effective problem solving for complex applications in a particular sector (e.g., healthcare, manufacturing, agriculture, etc.) identified by the Smart Specialisation Strategy of Cyprus (S^3Cy). Objective is to target emerging concerns that can be highly relevant to global development, support new waves of productivity growth and innovation, and improve people's lives.

The project partners have successfully translated and condensed their expertise in a Strategic Research & Innovation Agenda (SRIA). The SRIA captures ongoing developments and provides CUT researchers with important perspective and insights on future research priorities in this fast-paced technology field that can evolve into a decisive concept for solving some of the key challenges for humankind. The purpose of the SRIA is to actively reshape fragmented research at CUT into a cohesive, strengthened collaborative research-model. It aims to support the alignment and coordination of national and leading partner research that will achieve innovation, connectivity, research synergy, and stimulate human resource development. In this context, this deliverable addresses the research and innovation priorities for the future Smart Data Processing and Systems of Deep Insight technologies and applications that will drive changes across industrial sectors, the European economy and society in general.

The body of research work in the SRIA is topically organized in a set of three inherently inter-related broad research themes, called JRAs in deliverable D2.1 - Survey on Smart Data Processing and Systems of Deep Insight: Current Research and Future Challenges. The subject areas within the DESTINI JRAs (Smart Data Processing, Systems of Deep Insight, and Methodology for Smart Data-centric Services & Applications) emphasize research focus on long-term innovative research aimed at achieving breakthroughs by enhancing CUT's research capacity, potential and talent pool while addressing significant gaps and emerging issues with national implications in priority sectors identified by S^3Cy.

Several challenges and open research problems were reported and documented in deliverable D2.1. These were organised into classes that describe the nature of the challenge or problem as follows:

i. Optimization, Decision Support, Prediction
ii. Utilization of artificial or computational intelligence
iii. Enhancement of experimentation
iv. IoT infrastructure, technology and applications
v. Data modeling and characteristics
vi. Cloud environment
vii. Smart data processing
viii. Standardization

The SRIA of DESTINI is built taking into consideration the findings of the survey and focusing mostly on items (i), (ii), (iv), (v) (vi) and (viii). As the challenges above are intertwined, the rest of them are also considered but in a mote peripheral manner.

The rest of the document is structured as follows: Section 2 describes the Characteristics of Smart Data Processing and Systems of Deep Insight that are central to the SRIA, while section 3 outlines general concepts of the SRIA. Section 4 presents the research and innovation agenda of DESTINI organized in three major domains, smart healthcare, smart manufacturing and complementary priority axes. Section 5 describes the methodology for application development starting with general strategic areas and ending with focusing on the two major domains of interest; this section also outlines collaboration with strategic partners, stakeholders and industrial/market players. Section 7 presents some plans for further work if time permits or future steps. Finally, the deliverable closes with some concluding remarks.

2 Characteristics of Smart Data Processing and Systems of Deep Insight Central to the SRIA

Smart Data Processing and Systems of Deep Insight combine cognitive functions with sensing, data communication, integration and improved decision making in an integrated way. Figure-

1 illustrates the DESTINI knowledge areas and their interconnection. What differentiates a smart system from a system that is purely reactive is the reliance on smart data and analytics abilities, which gather and interpret information, and build a sound foundation for taking into account user perception, leading to actionable insights while providing the basis for actionable "intelligence" and a move toward more fact-based decisions.

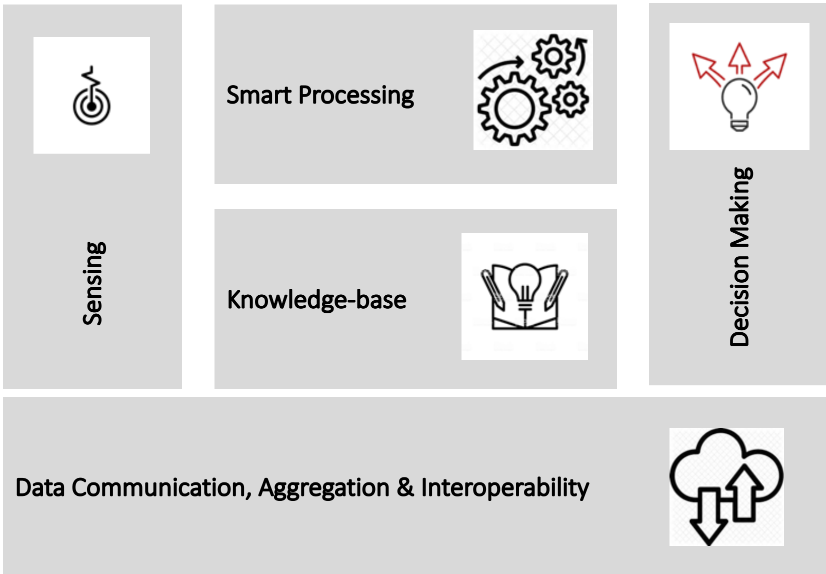

Figure 1. DESTINI knowledge areas & interconnection.

The area of Smart Data Processing and Systems of Deep Insight includes the ability to clearly define, interoperate, openly share, access, link, syndicate, transform and manage data. Under this perspective, it becomes crucial to have knowledge-based, meta-data representation techniques to structure the data sets and content, annotate them, link them with associated processes and software services, and deliver or syndicate information to recipients. These are the mechanisms that convert stale data to smart data. This area also includes adaptive frameworks and tool-suites in support of smart data processing by allowing the best use of both data in motion (e.g. data streams from IoT sensors), and data at rest, and may rely on advanced techniques for efficient resource management, and partitioning of intensive data workloads across a number of private and public clouds.

Smart Data Processing and Systems of Deep Insight (SDP&SDI) are often integrated with the (natural, built and social) environment, networks of Big Data, other smart systems and the human. It is a sufficient (or extrinsic) condition for the smartness of a system to provide (and use) cognitive support to (and from) its surroundings. System functionalities determine advancements in "smartness". These can be expressed in terms of smart systems that perform human-like perception and actions. Intrinsic to this approach and central to the SRIA activities described below are the concepts of smart data and digital twins.

We use the term *Smart Data* to emphasize the latent value inherent in widely dispersed and unconnected data sources. Aim is to create meaning from data, consider multiple scenarios and provide decision-makers with high value information and services that help them make the best possible decisions and solve complex problems. Smart data exhibits the following properties:

- Normalised Data – these are conflict-free, homogenised data retrieved from multiple related data sources and diverse representations that can be interpreted in a specific context.
- Contextualised Data – these are normalized data given meaning and contextual-awareness to enable orchestration and improved decision-making.
- Orchestrated Data – these are cross-correlated secure contextualized data across a specific domain (e.g., healthcare, manufacturing, smart cities, etc.) that can be turned with AI support to actionable tasks at the speed of business.

A *Digital Twin* is a digital model of a physical asset and/or process, or service, e.g., of a patient or novel product, which can be used as a virtual test-bed for improved decision making. This mirroring of the physical world and its pairing by a virtual image allow for deeper analysis of smart data that collectively characterize the physical world and monitoring of systems to head off problems before they even occur, develop new opportunities, predict future behaviour and plan for the future by using simulations. For instance, using a patient digital twin we may enable doctors and other healthcare providers to capture and track patient data in order to tailor treatment to each patient, reduce medical risks, and generate a more accurate therapy.

3 General Strategic Research & Innovation Agenda Concepts

The DESTINI SRIA revolves around transversal topics, which represent the SDP&SDI Joint Research Activities (JRAs), and application domains (which reflect application sectors that are part of the S^3Cy). The SRIA is conceived and generated to direct future research on SDP&SDI technologies and applications and to guide development and innovation in Cyprus and in Europe.

The partners have identified three Joint Research Activities (JRAs) for the DESTINI project:

- <u>Smart Data Processing Systems:</u> This JRA includes data ingestion, data aggregation of an enormous variety of structured, unstructured and semi-structured datasets, knowledge-based meta-data representation techniques for the conversion of raw into smart data, data privacy and protection, automated deployment, run-time software performance monitoring and dynamic configuration.
- <u>Systems of Deep Insight:</u> This JRA focuses on analytic solutions that enable optimization of asset performance in smart data processing systems and is geared towards systems of insight. These are systems that turn data into insights, systematically test insights and find those data that matter to make them contextual and actionable.
- <u>Methodology for Smart Data-centric Services & Applications:</u> This JRA targets smart application development techniques by providing a methodology that interlocks elements of smart data processing and systems of deep insight to alleviate complexity and the effect of changes, thus speeding up the entire software development/deployment process for smart applications in priority sectors identified by the S^3Cy.

The objectives, research strategy, and impact, as well as the research overview of each transversal topic, have been extensively described in deliverable D2.1.

We have identified for the SRIA the following two important application domains: Smart Health, and Smart Manufacturing/factory automation, which are part of the S^3Cy. These two domains have been identified due to their importance and relevance to Smart Data Processing and Systems of Deep Insight. In the original proposal we had considered the domain of shipping, but the partners decided that the medical domain is more timely, has greater impact and presents much more fertile ground for research. In addition, the leading partners have funded projects and experience in the medical domain which facilitates knowledge transfer to the widening institution.

In addition, we have identified the Internet of Things and Digital Twins as important contributing technologies, which are central to research activities that characterize Smart Data Processing and Systems of Deep Insight and are perfectly aligned with the challenges reported in deliverable 2.1. A detailed overview of the Application Domains and the Transversal Topics of the SRIA are shown in Figure-2. The transversal topics and application domain in the SRIA are described in terms of objectives, research strategy, and impact. These are described in the following.

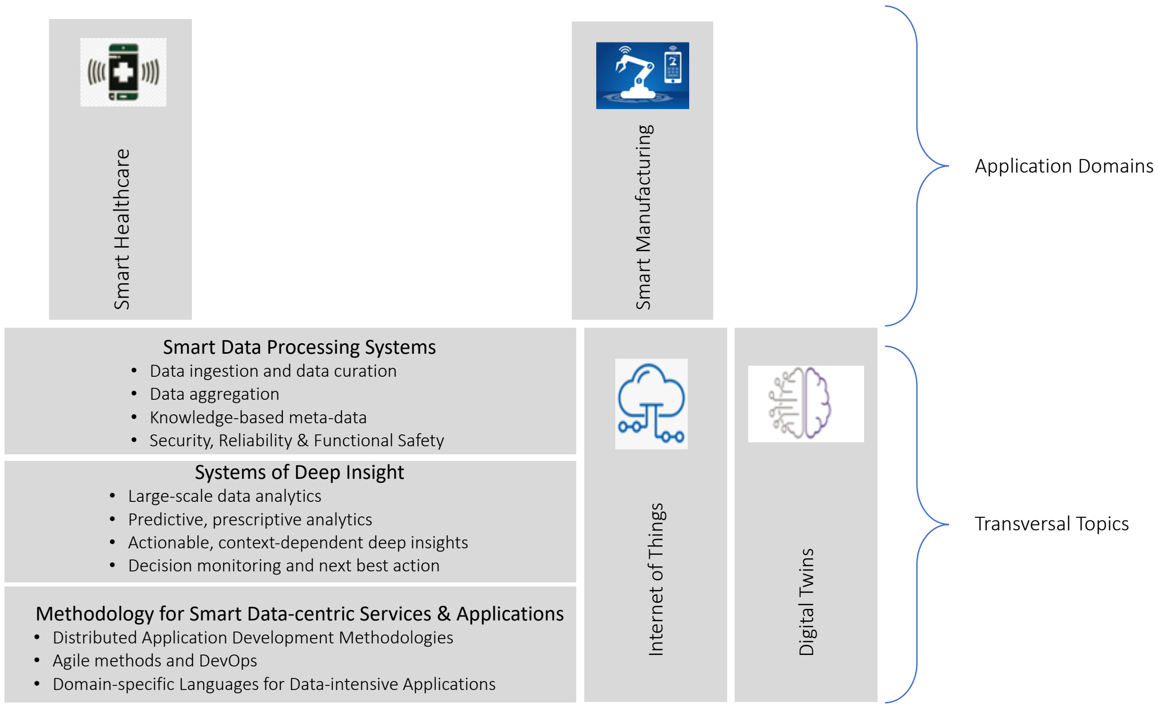

Figure 2. DESTINI SRIA

Challenge: describes general characteristics of the application domain, open problems and trends, as well as future directions.

Scope: describes the type of application problems that will be considered and how SDP&SDI technologies can help to solve these problems.

Objectives: After a short general definition of the application domain, links are made to the societal challenges this application domain affects, the roles of SDP&SDI to meet such challenges are explained, and the objectives of SDP&SDI regarding developments in the application field are described.

Expected Impact: The socioeconomic impact of the application field and the share of SDP&SDI therein are stated, and the links to challenges for the R&D and societal challenges and objectives are made.

Strategic research areas: describes the individual areas of research that contribute to application domains. Specific SDP&SDI technologies in each application field are analysed, and the research, development and innovation challenges regarding the objectives are stated.

Following the formulation of the application domain objectives, research strategy, and impact for each of the three application domains identified above, a *research road map* for the SRIA will be devised. The roadmap describes actions in terms of R&D for Smart Data Processing and Systems of Deep Insight needed to achieve the short and long-term objectives of the project. The roadmap is structured along milestones that show the project progress for the years to come. These are linked to defined objectives and strategies within the application domain and the transversal domains. The actions indicate what needs to be done and when in terms of SDP&SDI for each application domain and the enabling technologies (for transversal topics) to achieve the milestones.

4 SRIA and Application Domains

4.1 Smart HealthCare

The overall objective of this section is to provide the DESTINI vision for smart systems-based solutions to address smart healthcare challenges, as well as to satisfy clinician, care provider, and patient needs (Ullah et al., 2016).

Challenge: Today's predominant healthcare paradigm is "disease management," where providers wait for individuals to become ill and present with symptoms. Although this disease management model solves certain problems, there are still many inadequacies such as inefficiency and inability to synchronize patient and treatment data (Bardram, 2008). As the number and complexity of health conditions increase over time and episodes of acute illness are superimposed, the type and number of care providers contributing to the care of individuals also increases. It becomes significantly more difficult to align and coordinate care across care teams and associated settings. This results in fragmented care, due to poor communication and information sharing. Without secure information exchange among the actors involved in health, social and informal care services and a process to reconcile potentially conflicting treatment plans, it is impossible to avoid redundant and potentially harmful interventions.

Scope: Aim is to understand how to achieve an effective continuum of care, by combining and strengthening existing solutions, leveraging real-world evidence and centering solutions around the patient. Research work will investigate the concept of "precision health care" (e.g., disease detection and prevention) by proactively monitoring chronic desease patients, performing preventative and wellness interventions, and managing prevention and wellness for at-risk individuals.

Building on automation and Artificial Intelligence (AI) technologies (e.g., knowledge-based techniques, machine learning, predictive/prescriptive analytics, etc.), which bolster Smart Data Processing and Systems of Deep Insight, have the potential to revolutionise healthcare and help address some of the challenges set out above (Yuhanna, 2014). We shall investigate how precision health models underpinned by Smart Data Processing and Systems of Deep Insight (e.g., predictive analytics based on machine learning) and new carriers such as wearable and IoT devices, social media, and medical big-data platforms can record and analyse masses of personal health data, providing patients with new medical services for prognosis and preventive care management, as well as improved treatment (Jiang et al., 2017).

Objectives: Enabling a novel approach to smart healthcare to strengthen disease prevention and integrating service delivery around people's needs for health and social care that puts more emphasis on prevention and adopts a person-centred rather than a disease-centred approach. Aim is to focus on personalised early risk prediction models, estimating the probability that a specific event occurs in a given individual over a predefined time, to enable earlier and better intervention.

The challenge is to develop and validate these comprehensive models based on AI or other state of the art technologies for *prediction, prevention* and *intervention* using multiple available data resources, which are integrated in personalised health and care pathways that empower individuals (self-management) to actively contribute to risk mitigation, prevention and targeted intervention. To this end, the DESTINI SRIA aims to provide AI-infused precision and goal-directed healthcare focusing on prevention and well-being, identification of potential risks and on suggesting preventative interventions. More specifically, it will leverage analytics, machine learning methods and conversational interface technologies to provide the following functionality:

1. Enhance patient experience and boost their personal care by providing personalised interaction sessions and content (including potential customisation options for patients with comorbidities and complex diseases) and streamlining patient access to care;
2. Decide how to tune the interactive experience to the individual stakeholder (patient or clinician);
3. Mine the huge amounts of genomic, clinical, life-style, and real-time patient data to assist patients, as well as clinicians, by making appropriate recommendations, and personalize medical treatments;
4. Serve as a foundation to build successful digital medical chains and clinician and hospital partnerships and collaborations thereby improving healthcare productivity;
5. Accelerate the speed at which improved treatments are delivered, while reducing the cost in doing so; and,
6. Employ continuous digital monitoring to monitor the patient’s health by connection to patients’ wearable and IoT medical devices and notify both the patient and provider should the AI-enabled monitoring determine that intervention is required.

Given that chronic diseases may be caused by multiple underlying mechanisms in different patients, the AI platform will act as a virtual health assistant to provide a personalized experience in which patients can ask questions and learn how to better manage their health and supply personalized treatments in collaboration with medical professionals that will improve the patient's well-being and eventually cure the disease.

Expected impact: Smart Data Processing and Systems of Deep Insight have the potential to transform how care is delivered. They can support improvements in care outcomes, patient experience and access to healthcare services. In addition, through mobile healthcare, patients can better understand their own conditions and strengthen their contact with doctors, thus effectively improving the interaction and relationship between doctors and patients. SDP&SDI can add value and help improve the diagnosis process and access to treatment, leading to a precision and goal-directed healthcare that involves multi-level strategic planning for medical care. With this approach, the patient's goals are set first, and then their operationalisation is implemented into clinical practice.

The expected impact of this approach is substantial, is in consonance with the *hospital of the future* (Pickering et al., 2012) and includes:

- Predictive health versus currently applied reactive 'disease care' approach;
- New dimensions of medical interests and tasks;
- Highly effective cross-border and cross-professional cooperation;
- Active participation of patients in the healthcare processes; and,
- Better interoperability to facilitate greater collaboration and inspire new opportunities for concepts like telemedicine to help alleviate ever growing physician workload.

4.1.1 Strategic Research Areas that Contribute to Smart HealthCare

4.1.1.1 Smart Data Processing Systems and the Creation of a Medical Data Lake

Smart Data Processing supports the processing and integration of data into a unified view from disparate medical Big Data sources ranging from data warehouses, image data, sensors and devices in the Internet of Things, and social platforms to databases, whether structured or unstructured, to support Big Data analytics (Lee et al., 2017). Smart data processing in SRIA will support the processing and integration of data into a unified view from disparate Big Data sources, data warehouses, sensors and devices in the Internet of Things, and databases, whether on-premises or cloud, structured or unstructured, to support Big Data analytics. The aim is to select, aggregate, standardize, analyse, and deliver smart data to the point of care in an intuitive and meaningful format. To achieve this, the following activities could be pursued as part of the *SRIA roadmap milestones* (in conformance with objectives in section 4.1):

1. Knowledge-based metadata representation techniques to structure disparate medical data sets and content, annotate them, link them with associated processes and

software services, and deliver or syndicate information to recipients. These are the mechanisms that convert stale data to smart data.

2. Systematic medical representation and interoperability language: an expressive meta-data language that will convert medical data into medical knowledge structures by transforming and integrating heterogeneous medical data sources to allow seamless interoperability, in a manner that surmounts vocabulary, structure, and format data inconsistencies.
3. Creation of a cohesive Medical Data Lake from structurally and semantically enhanced data collection to provide the basis for actionable insights on emerging concerns that can be highly relevant to improving QoL.
4. Data in the Medical Data Lake can be organized and made accessible on a when needed basis and subsequently made actionable for analytics using late-binding techniques.

A *Medical Data Lake*, as the name implies, is an open reservoir for the vast amount of data inherent within healthcare, e.g., traditional sources of health data (cohorts, comprehensive electronic health records or clinical registries, including genetic data, validated biomarkers for remission), from new sources of health data (mobile health apps and wearables) and sources that are usually created for other purposes such as environmental data, which can be integrated into an analytics platform to improve decision making. A data lake can ensure that data can employ data security and privacy mechanisms to ensure confidentiality and anonymity of data transfer to avoid misinterpretation and inappropriate conclusions by using proper annotation methodologies of the data.

The above SRIA milestones are structured in chronological sequence. Milestones 1 to 3 are short- to medium-term, which span the duration of the project. Long term SRIA milestone 4 could be pursued only if additional partners are added to DESTINI through its networking activities with EU funded or domestic projects of similar nature.

4.1.1.2 Deep Medical Insights and Informed Action Taking

The DESTINI SRIA introduces a separation between knowledge representation formalisms (in the subsection above) and inference (in the current subsection). While Smart Data Processing Systems technologies are used to develop expressive representation and interoperability tools, Systems of Deep Insight take advantage of the timely, accurate, and smart data (developed by Smart Data Processing Systems above) to support informed decisions and resultant actions of clinicians and patients, and produce better outcomes across the entire continuum of end-to-end healthcare. The SRIA connects machine learning algorithms to smart medical data, bringing a variety of AI-driven functionalities to clinicians and patients (Dimitrov, 2016). Analytics will primarily function as decision-making aides by gathering and interpreting information and building a sound foundation for clinical decision-making.

The SRIA will employ techniques of Deep Medical Insight – based on AI analytics – to develop the following activities which could be pursued as part of the *SRIA roadmap milestones* (in conformance with objectives in section 4.1):

1. Discover (hidden) associations between smart data, prioritize results, find useful insights, and discover large-scale patterns and trends within the smart data to reveal a wider picture that is more relevant to patient care.
2. *Manage knowledge in networked healthcare organizations* from the above step to aid clinical decision-making by transforming information into personalized patient actionable intelligence that can be interpreted by both clinicians and patients. DESTINI will employ machine learning mechanisms, which can reuse past results to inform its next predictions and lead to truly actionable clinical intelligence.
3. *Explore ways of inducing adequate personalised preventive measures* (e.g. behavioural change, diet, interventions, medication, primary prevention) from *advanced predictive models*. More specifically, this step aims at creating personalized interactions that meet patient needs when under treatment or follow-up, allowing providers to collect and send key-data to improve the long-term management process.
4. *Develop shared interpretation approaches* that aim at fostering shared meanings of what is happening outside and inside the healthcare organization in order to plan and make decisions. Shared interpretations are needed to define the organization intent or vision about what new knowledge and capabilities the organization needs to effectively monitor the health status of individual patients, provide overall actionable insights at the point of care, and improve quality of life after treatment.
5. *Understand patient cultural and personal preferences* by combining patient-reported outcome measures (PROMS) and patient-reported experience measures (PREM) data to present patients' perceptions of both the process and outcome of their personalized care. Such data not only help improving and focus patient-centred clinical management, but also provide vital feedback to health care providers to allow comparisons in clinical care.

The above SRIA milestones are structured in chronological sequence. Milestones 1 to 3 are short- to medium-term, which span the duration of the project. Long-term SRIA milestones 4 & 5 could be pursued only if additional partners are added to DESTINI through its networking activities with EU funded or domestic projects of similar nature.

4.2 *Smart Manufacturing*

The overall objective of this section is to provide the DESTINI vision for smart systems-based solutions to address smart manufacturing challenges revolving around the need to produce innovative customized products. It focuses on the ability not just to collect (large) amounts of manufacturing data, but also to effectively and efficiently analyse it to learn previously unknown insights in the product customisation and production processes (Tao et al., 2018).

Challenge: In the age of Smart Manufacturing (or Industry 4.0), there is a need for an open, innovative, and customer-focused product development culture (Lasi et al., 2014). Digital development needs to captivate the growing expectations of today's customers for product customisation, introducing novelty, and getting products to the market faster. The emergence of digital products has reintroduced the concept of product customisation at a wider scale. Creating innovative products that customers demand requires manufacturers to develop an outside-in, customer-centric process that can react quickly to changing needs and incorporate customer and product designer feedback. It also requires designing product parts, adjust processes and production lines for ease of manufacturing of end-products with an end goal of making a better product at a lower cost.

Scope: aim is to understand how creating innovative products on-demand can be achieved by optimizing, refining, and making product design and development nimbler to react to customer expectations and market pressure. This is where SDP&SDI technologies are coming handy as they can demonstrate that they can provide help in automating product design and transform customer experiences as well as a much-needed productivity boost.

For product design and development, SDP&SDI comes into play via a new process called Generative Design. Generative design enables a design exploration process in which designs are no longer generated directly by designers but is assisted by smart data and machine learning algorithms. Designers or engineers input design goals and parameters such as performance or spatial requirements, material type, manufacturing techniques, resilience, cost constraints and more into the generative design software. The designer or engineer evaluates the generated design alternatives and modifies the settings for designs that do not work, while the software learns from each iteration.

Objectives: Customisation in manufacturing currently involves minimal flexibility to alter only fixed parameters. To achieve advanced customisation, new modelling and processing languages for representing, exchanging and processing product and production knowledge that facilitate the transition from digital to physical products are necessary.

To achieve its mission and objectives, the SRIA requires integrating innovative and disruptive technologies and capabilities geared towards customisable products that have smarter, dependable, and secure plug-and-play integration of digital and physical components.

The SRIA aims to provide a comprehensive fusion of generative design with digital twins by providing a richer model of a customized digital products, ranging from ideation and design through product realization and utilization. This enriched model directly results in a holistic description of all facets of an innovative digital product development and all of its constituents, its performance expectations, and the manufacturing process that will eventually construct the product under real-world conditions.

Expected impact: Generative design tools support agile production processes and predictive quality techniques (Dubey et al., 2015) that allow product designers to pursue product customization goals and constraints by generating optimized customized product features. These features meet customer/product designer objectives and domain-specific requirements in terms of product features, quality, time criticality, safety and security.
Combining generative design tools with manufacturing digital twins is the catalyst behind digital reinvention. SDP&SDI helps create customized products that evolve with their users and continually enhance the customer experience. Industrial incumbents that can establish such digitally reinvented products in the market at scale will dramatically boost top-line growth, and thus their market capitalization. Product developers and manufacturers in several dimensions including:

1. Shortening production ease and cycle for customizing a new product;
2. Reduction of time to produce new customizable products;
3. Reduction of overall cost to produce new customizable products;
4. Improvement of decision making regarding product customization/differentiation;
5. Improvement of quality of innovative new products;
6. Increased customer satisfaction;
7. Increased efficiency, flexibility and robustness of the customized product and production process is also expected.

4.2.1 Strategic Research Areas that Contribute to Smart Manufacturing

4.2.1.1 Smart Data Processing Systems for Smart Manufacturing

Smart Data Processing supports the processing and integration of data into a unified view from disparate manufacturing Big Data sources ranging from data warehouses, image data, sensors and devices in the Internet of Things, and social platforms to databases, whether structured or unstructured, to support Big Data analytics (Russom, 2011). This includes critical product and process data, data used to ensure/assert quality and reliability of manufactured parts, data used to drive production processes, decision-critical operational data utilized for offline training of machine learning models and run-time inferencing.

Digital product customization in the SRIA will involve end-users configuring a standardized product by adding, removing or replacing product parts. Product configuring comprises baseline product variants that share the same parametric description and a variant that can be generated by scaling one or more parameters.

To achieve the above, the following activities could be pursued as part of the *SRIA roadmap milestones* (in conformance with objectives in section 4.2)

1. Improving the data assets by addressing data pre-processing challenges for the various product and production data types (including unstructured data such as image, text, etc. and real-time data) and processes to improve interoperability and reveal a wider picture that is more relevant to digital product customisation. This includes methods for annotation of unstructured data sources, unbiased and representative input data, methods for handling volumes of real-time data with high velocity, etc.
2. The development of data augmentation methods for transforming data assets into high-quality smart data that makes the sensing, measurement and perception of manufacturing product data assets amenable to use in decision-making. By combining data-driven and knowledge-based models, it becomes possible to:
   - close the loop from data-driven, automated analytics and decision support to fully automated enactment and actuation of decision, a significantly higher level of automation and reliability of production processes becomes possible.
   - have a sustainable digital-twin along the complete lifecycle (product and production) that provides value to AI data integration.
3. Methods for knowledge modelling and representation that enable the seamless integration of data and connection with the physical world. To support reuse of integrated and continuous knowledge its representation in a standardised format.
4. Combine components to form a final product using product configuration techniques that analyse existing components to verify that they are efficiently designed, they fit within a novel product that can be manufactured.
   - Employ modularised product customisation mechanisms for Smart Data Processing systems which target of plug-and-play solutions and clearly described (semantic) functionalities for digital products that can be further reused and upgraded.
5. Generating domain related knowledge representations and language establishing the basis for seamless incorporation of background knowledge into AI applications. This may include approaches that combine data-driven learning with symbolic approaches (hybrid AI), digital twins, simulation technologies and methods that enable the data processing at the location where the data is produced (edge analytics) and methods for knowledge representation learning.

The above SRIA milestones are structured in chronological sequence. Milestones 1 to 4 are short to medium term, which span the duration of the project. Milestone 5 includes medium term and long-term activities. Long term SRIA milestone 5 activities (such as simulation and edge analytics functionality) could be pursued only if additional partners are added to DESTINI through its networking activities with EU funded or domestic projects of similar nature.

4.2.1.2 Deep Manufacturing Insights and Informed Action Taking

The SRIA will focus on the concept of generative design-based digital twins by tasking an AI-based generative design and configuration approach and algorithms with exploring the design space, evaluating alternative customization options and then reporting back to the designer which options it considers promising for further analysis. Such a system allows a much deeper exploration of complex design spaces. Thus, changes become faster and more natural to handle. Product configuration will serve as the channel that surrounds all activities that contribute to synthesizing customized products from a set of predefined components, be they parts or assemblies, while respecting a set of well-defined constraints, which restrict how product components can be selected and combined to create a customized product.

Systems of Deep Insight in the SRIA will be involved in all aspects of the manufacturing process, providing “intelligent design” in developing products, supporting effective customization decisions and planning, ensuring product consistency, recommending performance improving actions, and making products more useful and more reliable (Dagli, 2011). That, in turn, can bridge the gap between physical and digital worlds, creating new products and services and evolving these into customer outcomes.

To achieve the above, the following activities could be pursued as part of the *SRIA roadmap milestones* (in conformance with objectives in section 4.2)

1. In the conceptual phase of product development, Machine Learning and Systems of Deep Insight can be applied in combination with virtual engineering models and simulation for iterative design. With these technologies, numerous design options can be cycled through instantly, with recommendations automatically generated for optimal solutions based on multiple criteria (e.g., structure, cost, sustainability, time, regulatory requirements, etc.). To this end, a rigorous Generative-based Digital Twin model can be developed to optimize designs, prepare a digital product for manufacturing and define the total path for production.
2. Develop human-centred manufacturing solutions including human-machine relations, interaction, collaboration, and complementarity. Human-in-the-loop includes human-

as-part-of-the-manufacturing- system, including intuitive systems, human–machine collaboration and collaborative decision making.

- Allow the designers and lay users to employ model-driven manufacturing techniques based on AI and deep learning to display, analyse and modify the Digital-side Twin model resulting innovative, functionally stable customized products.

3. Deep insight-powered digital modeling and simulation (including virtual reality systems) are also being used to plan production lines and systems, monitor product quality and production performance and funnel important production data back to teams working on product design and specifications.
4. Develop deep learning mechanisms and involve IoT, big data, machine learning and AI, in advanced production monitoring, which can be supported by modelling and simulation/digital twins. Key-priority areas include (automated) modelling and analytics mechanisms and multi-variant and multi objective-simulation and optimisation.

The above SRIA milestones are structured in chronological sequence. Milestones 1 to 3 are short- to medium-term, which span the duration of the project. Milestone 4 includes medium-term and long-term activities. Long-term SRIA milestone 4 activities (such as multi-variant and multi objective-simulation and optimisation) could be pursued only if additional partners are added to DESTINI through its networking activities with EU funded or domestic projects of similar nature.

4.3 Additional Priority Research Areas

The SRIA exercise has identified by the partners priority research areas that need to be investigated in the context of transversal topics and application domains. The importance and linkage of these technologies with elements of the SRIA as shown in Figure-2 are briefly outlined below.

4.3.1 IoT Technologies in Support of Application Domains

IoT represents a network made up by billions of heterogeneous physical devices, such as smart meters, smart locks or wearables, connected to the Internet. These devices collect and share data about their surrounding environment to favour automation and optimisation of different tasks related both to people's daily life and to industry.

The first priorities for IoT research and innovation in the next years are in the areas of IoT distributed architectures, edge computing, end-to-end security, Blockchain, AI and the convergence of these technologies. IoT and edge computing will see innovation and wide adoption in both consumer and industrial IoT, enabling better security practices and reducing connectivity costs (Shi et al., 2016). In this context, the convergence of connectivity, IoT, edge computing, AI, and Blockchain will be essential to next-generation Smart Data Processing and Systems of Deep Insight applications and advancements.

IoT technology coupled with AI technologies that underpin SDP&SDI can provide a foundation for improved and, eventually, entirely new products and services. The powerful combination of AI and IoT technology brings new challenges in addressing distributed IoT architectures and decentralised security mechanisms.

In the future of IoT applications, it is expected that AI techniques and methods are increasingly embedded within several IoT architectural layers and create artificial intelligence of things applications. The AI segment is currently fragmented, characterized with most implementations focusing on a silo approaches to solutions. Future trends include solutions involving multiple AI types and integration with IoT and possibly Blockchain.

The IoT is commonly known in the healthcare industry as the Internet of Medical Things (IoMT) consists of any and all medical devices, patient monitoring tools, wearables, and other sensors that can send signals to other devices via the Internet. The combination of Artificial Intelligence (including Machine Learning, Deep Learning, Natural Language Processing, etc.) and the Internet of Things will support the development of smart hospitals and fuel the ongoing growth of big data analytics and spark a new era in smart healthcare (Calo et al., 2017).

Major challenges and open research problems that could be researched include:

- <u>Patient self-diagnosis:</u> Using IoT wearable tools to monitor your health and even carry out patient own scans, patients will finally have the ability to self-diagnose a wide number of conditions at home, without needing to visit a surgery or hospital. This will be beneficial to enable multi-faceted telehealth services by making use of affordable devices for interfacing medical practitioners, clinicians and patients together. It also proves useful for data transmission through connected devices like monitors, pacemakers, and other devices.
- <u>Patient management perspective:</u> SDP&SDI is ideally placed to drive preventative care to keep patients well in and out of the hospital system. By having sensors and support tools in people's homes and feeding into analytics and AI, we can monitor their health

remotely. On the analytics side, we have a push towards machine learning, allowing clinicians to act and take actions to deliver preventative and proactive care wherever and whenever needed to make the medical system sustainable."

In the domain of Smart Manufacturing, IoT provides a digital infrastructure for communication, integration, interoperability, human-machine interaction and remote operations, enabling a huge ecosystem of cloud services, also based on data analytics and/or AI.

Major challenges and open problems that could be researched include:

- Automating both the monitoring of systems across production lines and the factory floor: Ingesting massive data from heterogeneous sensor embedded on machines, environment and infrastructures enable monitoring of working parameters to ensure a detection before failures occur during production (e.g., when defective materials have found their way onto the production line) and automatic remediation of certain problems proactively.
- Improving production planning and scheduling to eliminate manual inspections from production processes that may consume up to 30 percent of the total production time. To automate checks in order to increase throughput, IoT sensors can be installed to monitor temperature, humidity and dust levels throughout the manufacturing process and an analytics solution can then ingest data from the sensors to deliver real-time insights into changes that might impact the quality of products and components being produced.

4.3.2 Digital-Twin Technologies in Support of Application Domains

By creating a digital-twin, insights about how to improve operations, increase efficiency or discover an issue are all possible before it happens to whatever it is duplicating in the real-world (Tao et al., 2018). The lessons learned from the digital-twin can then be applied to the original system with much less risk and a lot more return on investment. SDP&SDI can be used to analyse the model of operations represented by the digital-twin no matter where the physical object is located. Below we examine the use of digital-twins in the application domains in the SRIA.

A *Medical Digital-Twin* can be defined, fundamentally, as an evolving digital profile (personal patient digital replica) that captures the historical and current behaviour of a patient, tracks all sorts of patient data and associated medical process and treatment. Overall, it helps optimize patient health and can be used as a virtual test-bed for improved future treatment. The Medical Digital-Twin has as a purpose to:

- enable doctors and other healthcare providers to capture, associate and track patient data in order to tailor treatment to each patient;
- incorporate a variety of care data, including vital medical information from medical records, current medication, imaging studies, lifestyle, genetic, and patient-provided health data from exercise or health monitoring applications and medical pathways;
- reduce medical risks and generate more accurate therapy for patients.

A *Manufacturing Digital-Twin* can be defined, fundamentally, as an evolving digital profile of the historical and current behaviour of a physical object (product or machine) or process that helps optimize performance. The digital-twin is based on massive, cumulative, real-time, real-world smart data measurements across an array of dimensions. These measurements can create an evolving profile of the object or process in the digital world that may provide important insights on system performance, leading to actions in the physical world. With the creation of the digital-twin, companies may realize significant value in the areas of speed to market with a new product, improved operations, reduced defects, and emerging new business models to drive revenue.

The ultimate objective of artificial intelligence and machine learning is to enable the development of a digital-twin of the product and production processes. The creation of a digital-twin should take place under a model-based systems engineering process using machine learning algorithms and knowledge gained as a foundation. The digital-twin can serve as a platform for running what-if scenarios to improve product customisation when used as a model for designing higher reliability parts and adjusting the interactions between production-line machines to improve performance.
The Manufacturing Digital-Twin has as a purpose to:

- Blend generative design techniques with the concept of Digital Twin to create an “alive entity” that continuously learns and updates itself from multiple product sources and past experiences (e.g., historical product designs and past client requirements).
- Facilitate collaboration between humans and AI-based systems by developing a user-friendly interactive, digital-twin experience wherein a product generative design is created where product parameters and constraints (including safety and security) can be specified by users, changed, programmed and managed creating a wealth of interlinked data and knowledge structures that drive production workflows.

5 Methodology for Application Development

5.1 Strategic Areas

Artificial intelligence (including Machine Learning and Deep Learning), which underpins Smart Data Processing and Systems of Deep Insight, is gaining widespread grip in a number of application domains. The types of applications developed by SDP&SDI underpinned by AI can be characterised as large-scale smart applications. Large smart applications can be broadly thought of as involving distributed data-intensive software development in conjunction with effective AI analytics and machine learning techniques that utilize critical and cross-correlated data produced in the course of operations. Data analysis challenges relate to processing and generating insights from the massive amount of data stored in diverse repositories. An additional characteristic is that the software development process in such smart applications started to be considered a socio-technical arrangement, where organizational and human aspects play a key-role and have to be supported by technology.

AI and machine learning frameworks in SDP&SDI comprise libraries of mathematical expressions and functions for various machine learning and deep learning operations. They often include a broad base of APIs and other development tools that are designed to assist developers in integrating into previous code and capitalizing on application systems for the data needed to train models and produce the application.

In the following we discern the most pertinent approaches to application development for Smart Data Processing and Systems of Deep Insight centric applications in the course of the project. All activities described below are medium- to long-term.

5.1.1 Digital Transformation

IT analyst firm Gartner Inc., defines “digital business transformation” as “the process of exploiting digital technologies and supporting capabilities to create a robust new digital business model” (www.gartner.com/en/information-technology/insights/digitalization). It is an important definition in that it places “business” in the middle of “digital transformation,” and it foregrounds the development of a new digital business model that is data-centred. Specifically, digital transformation involves using data, analytics, and connectivity to rethink everything product designers or designers of medical applications need to accomplish from the perspective of their customer (or patient / clinician / health carer), and using this analysis as the backbone for the development of new products and services for a compelling, personalized customer experience.

Digital transformations connect and enable analysis of every piece of data across channels, operation, and stakeholder outreach. For instance, digital transformation for smart healthcare applications enables the continual build-out and extension of services and data, while coordinating services across the care continuum to support patient centred health. Applications can range from providing personalized care options to gathering insights to addressing new care formats, such as telemedicine and outpatient care. In the following we shall focus on using healthcare examples to highlight important SRIA activities in the context of digital transformation.

In the medical domain, these SRIA activities may include:

- Developing a digital dexterity program and digital data-centric transformation strategy to enable a digital medical ecosystem to collaborate across organizational boundaries and unlock the flow of health data across borders (including clinical approaches, success stories, and experiences) and influence the quantity and quality of provided care;
- promoting a strategy for better alignment, normalization, aggregation, and exchange of medical data across medical agencies and borders and an integrated form of support and medical service provision;
- establishing a governance structure so that data is formally, consistently, and securely managed across an expanding number of healthcare providers and clinics and facilitates a better interaction and coordinated collaborative care among them.

Analogous activities can also be defined for the manufacturing domain.

5.1.2 Continuous delivery for machine learning (CD4ML)

Machine Learning (ML) applications are becoming increasingly popular, however the process for developing, deploying, and continuously improving them is more complex compared to more traditional software, such as a web service or a mobile application. They are subject to change in three axes: the code (business needs), the model (algorithms), and the data (schema). All these artefacts of the ML software production process require different tools and workflows that must be versioned and managed accordingly.

With an increased popularity of ML-based applications, and the technical complexity involved in building them, continuous delivery for machine learning (CD4ML) starts to gain traction in delivering reproducible and reliable ML-based application software releases, quickly and in a sustainable manner.

CD4ML is the discipline of bringing continuous delivery principles and practices to automate the end-to-end process of deploying ML applications into production. It removes long cycle times between training models and deploying them to production and manual handoffs between different teams, data engineers, and ML engineers in the end-to-end process of build and deployment of a model served by an application. Using CD4ML, developers can successfully implement the automated versioning, testing and deployment of all components of ML-based applications: data, model and code.

An important area of research for the methodological approach to application development in the SRIA is machine learning pipelines. A machine learning pipeline is used to help automate machine learning workflows in smart healthcare and smart manufacturing applications. They operate by enabling a sequence of data to be transformed and correlated together in a model that can be tested and evaluated to achieve an outcome, whether positive or negative.

ML pipelines consist of several steps to train a model. Machine learning pipelines are iterative as every step is repeated to continuously improve the accuracy of the model and achieve a successful algorithm. To build better machine learning models, for the purposes of CD4ML, we treat the ML pipeline as the final automated implementation of the chosen model training process. Once the data is available, one can move into the iterative workflow of model building. This usually involves splitting the data into a training set and a validation set, trying different combinations of algorithms, and tuning their parameters. That produces a model that can be evaluated against the validation set, to assess the quality of its predictions. The step-by-step of this model's training process becomes the machine learning pipeline.

5.1.3 Machine Learning Operations (MLOps)

Deploying ML models and applications in production settings where real-world decisions are being made by AI needs robust development, testing and staging cycles. MLOps (a compound of Machine Learning and "IT OPerationS") is a new practice for putting machine learning into production. MLOps applies DevOps principles to ML systems and applications. In its purest form, MLOps is the true instantiation of the automated production ML lifecycle.

Practicing MLOps means that developers advocate for automation and monitoring at all steps of ML application construction, including integration, testing, releasing, deployment and infrastructure management. The typical ML Deployment Lifecycle that can be considered for the SDP&SDI application development methodology may include the following steps (Sridhar, 2018):

1. Data preparation stage: This is the first step towards constructing an ML model. Most analytics or ML is only performed once data has been organized, structured and made accessible to data scientists, engineers and modelers. This involves the creation schemas and structure around data.
2. Creation of an ML model specific to a business need: Once the data has been readied, the next step typically involves developing ML applications that read this data and create a model specific to a business need (for example, a medical recommendation system for chronic disease patients would need a recommendation model). This stage normally involves multiple iterations on algorithm choices and tweaks, datasets, feature engineering, and evaluation.
3. Testing process: This takes the model through various combinations of datasets and configuration parameters. In order to prepare the ML application to be production-ready, this stage also involves a test strategy that involves scale and robustness. This process is typically performed with one or more business needs kept in mind. In case a model does not clear certain criteria, the training phase needs to be revisited.
4. Model deployment: The final stage involves deploying this model in production. Once out in the field, developers usually deploy analytics or reporting applications that track the impacts on the end-goals. Feedback gathered during this process can be used to impact the training phase again.

5.1.4 Automated Machine Learning (AutoML)

Automated machine learning (AutoML) is the process of automating the process of applying machine learning to real-world problems (Hutter et al., 2019). AutoML is the process of automating the time consuming, iterative tasks of machine learning model development. Essentially, AutoML focuses on automating repetitive tasks of the Machine Learning process.

AutoML covers the complete pipeline from the raw dataset to the deployable machine learning model. The high degree of automation in AutoML allows non-experts to automate the end-to-end process of model selection and training. Automated machine learning makes it possible for applications in every domain – healthcare, financial markets, the public sector, manufacturing, and more – to automate most of the modelling tasks necessary in order to develop and deploy machine learning solutions with ease.

AutoML tools for this user group usually offer a simple point-and-click interface for loading data and building ML models. Most AutoML tools focus on model building rather than automating an entire, specific business function, such as customer analytics or marketing analytics. However, most AutoML tools do not address the problem of data selection, data unification, feature engineering, and continuous data preparation.

Once a business problem is defined and the ML models are built, it is possible to automate entire business processes in some cases. It requires appropriate feature engineering and pre-processing of the data. AutoML is used to tackle classification and regression problems. Classification is a type of supervised learning in which models learn using training data, and apply those learnings to new data. Similar to classification, regression tasks are also a common supervised learning task. While in classification predicted output values are categorical, regression models predict numerical output values based on independent predictors. In regression, the objective is to help establish the relationship among independent predictor variables by estimating how one variable impacts the others. For example, a patient disease progression based on parameters like, blood pressure, glycose levels, unintended weight loss, fatigue, etc.

5.2 Application Development for Smart Healthcare

Medical professionals (e.g., doctors, nurses, surgeons and medical staff in general) do not have the required skills in machine learning, nor in software coding to build predictive models (Anderson et al., 2003). The medical data sets collected by these professionals typically access a number of patients and information such as age, gender, blood pressure, cholesterol level, blood sugar level, electrocardiogram results, average heart rate, maximum heart rate, etc.

The use of modern application development techniques based on SDP&SDI – such as the ones we identified above - allows medical health professionals to build predictive models from their data sets automatically. It is consequently possible to develop applications that predict if a patient is likely to have a heart attack knowing his/her age, gender, blood pressure, cholesterol level, maximum heart rate. These professionals do not need any specific training to use. Raw data can be used directly without any form of pre-processing: no normalization, no handling off outliers or feature engineering are required. The predictive results from processing these specific datasets can be obtained fairly easily. They can, for instance, generate a pertinent prediction which can help medics estimate the chance of a patient having a heart attack or a stroke.

5.3 Application Development for Smart Manufacturing

Predictive modelling to anticipate equipment downtime is referred to as failure prediction in manufacturing. These models are based on data collected from past failures of a given equipment (or similar ones). Machine learning is well suited to model current equipment behaviour and its potential breakdowns. Production equipment failures can be anticipated and maintenance can be scheduled before the problem happens, avoiding unnecessary costs.

Manufacturing, Maintenance and Operation Managers can benefit from predictive models (Anderson et al., 2003) using AutoML. They do not have the required skills in machine learning or coding experience to develop them from scratch. They collect, in the course of their daily activities, considerable amounts of data as most machines are equipped with sensors. Data such as temperature, pressure, moisture, exposure to light, duration of use since the last downtime, are typically collected and can be put to use by means of AutoML models.

The use of modern application development techniques based on SDP&SDI allows domain experts such as manufacturing managers, maintenance managers, operation managers, facility managers, to automatically build predictive models using MLOps. They can use their raw data directly: no normalization, no need to handle outliers or engineer new features. Thanks to this limited data preparation, the predictive results from an historical dataset can be obtained easily.

5.4 Collaboration with strategic partners, stakeholders and industrial/market players

The partners in DESTINI plan to pursue the establishment of collaboration links with strategic research and/or academic partners and organisations worldwide that may, on one hand further support and enhance their efforts to achieve research and scientific excellence, and on the other form working groups in the research areas of interest or join consortia to attract new funding in the future. In addition, DESINI will target at engaging businesses, practitioners, industrial & market stakeholders to prepare the ground for future real-world experimentation and validation. This will allow to enhance the knowledge that will be acquired either through this collaboration and/or by collecting valuable, real-world feedback from our stakeholders, using real-world feedback and most importantly real datasets. Several stakeholders have been contacted prior to submitting the proposal and have expressed their interest to participate either in the Advisory Committee of DESTINI or in the pool of stakeholders/collaborators depending on the category they belong to. More specifically DESTINI plans to engage the following stakeholders in Cyprus:

A. Public/governmental bodies
   (i) Ministry of Energy, Commerce, Industry and Tourism, Republic of Cyprus
   (ii) Cyprus Hotel Association
   (iii) Cypriot Ministry of Agriculture, Rural Development and Environment – Veterinary Services

B. Stakeholders List for preliminary piloting activities (priority areas in S^3Cy given in parentheses)
   (i) Paradisiotis Ltd. (Agriculture/Food Industry)
   (ii) Lemissoler Navigation Co. (Transport/Marine)
   (iii) Uniteam Marine Navigation Ltd. (Transport/Marine)
   (iv) Muskita Aluminium Industries (Structured Environment/Construction Industry)
   (v) Logisoft Computer Systems (Information and Communication Technologies)
   (vi) Lefkonoitziatis Dairies (Agriculture/Food Industry)
C. Other collaborators
   (i) Climate-KIC and the Innovation & Technology Office (INTENT)

Similar attempts to form a collaboration basis in the other two participating countries will be made targeting at complementing the aforementioned disciplines, market sectors and lines of businesses with local stakeholders in Italy and the Netherlands.

DESTINI will seek collaborations with selected projects in each of the target areas of application: (i) Smart Healthcare: H2020 project QUALITOP - Monitoring multidimensional aspects of QUAlity of Life after cancer ImmunoTherapy - an Open smart digital Platform for personalized prevention and patient management (https://cordis.europa.eu/project/id/875171) for better better monitoring of cancer immunotherapy patients, and, (ii) Smart Manufacturing: H2020 project ICP4Life - An Integrated Collaborative Platform for Managing the Product Service Engineering Life Cycle (icp4life.eu) for capturing novel digital product requirements and product customization needs. In addition, DESTINI will collaborate with two other ongoing Teaming projects in Cyprus, namely RISE - Research Center on Interactive media, Smart systems and Emerging Technologies (https://www.rise.org.cy/en-gb/) and EXCELSIOR - Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment (https://excelsior2020.eu/). The former will allow extending knowledge on digital-twins and embarking on advanced knowledge in advanced graphics (e.g. serious games) and virtual/augmented reality. The latter will open new perspectives as regards datasets and application of smart data processing to a set of challenging real-world problems, such as space data, satellite surveillance and remote sensing.

DESTINI will capitalize the potential to embark on International Research Collaboration Partnerships (IRCPs) in the near future to enhance knowledge development and the diffusion of new technologies and address, more efficiently, global challenges. IRCPs will be based on strong partnership and collaboration between interconnected universities, enterprises and administration. IRCPs can take different formats, including: international co-invention and coauthorship, licensing, pooling of resources, co-development of inventions, mobility of researchers, open access to research data and networks, applying for joint research project funds, institutes and facilities, and attracting international direct investment.

The bolstering of IRCPs with strategic alliances with influential industry partners and administration will have a spiral effect on the local society and economy. This will lead to improved possibilities through reputation and name recognition for the Centre to seek and obtain competitive funding in national and international fora, such as the Research and Innovation Foundation (formerly known as Research Promotion Foundation) and EU H2020 Programmes, as well as trigger RTDI investments at local and regional level.

DESTINI will target the creation of the appropriate conditions, both in scientific capacity and ability to form and document applied research ideas so that the submission of proposals in EU Framework and national programmes is promoted. Specifically, a number of EU calls will be addressed (e.g. Marie Curie, FET, FoF, Interreg, CSA actions, Erasmus+/Erasmus Mundus Master Degrees), as well as national calls of the Cyprus Research Promotion foundation and the Cypriot Ministry of Energy, Commerce, Industry and Tourism, and groundwork will be performed to investigate opportunities for funding, synergies with strategic partners and support from local industrial and market players. Additionally, a target is set for exploiting the stakeholders basis and discussing with key players for preparing industrial proposals and prototypes to attract funding from the private sector in the smart specialization areas of S^3Cy.

6 Future Work

As previously mentioned in section 3, the shipping sector was initially selected in the proposal to be one of the two pilots for studying real-world problems. More specifically, this pilot would have involved optimizing the route, refueling, and cargo weight definition of ships per route (harbor-to-harbor) so that CO_2 and Sulfur emissions are lowered, costs are minimized and the time of service is taken into consideration. As shipping data are kept for the routes, we planned to investigate how it may be coupled with on-line, open, real-time feed information (e.g. weather conditions) to assess the effectiveness of smart data processing methods and algorithms for achieving the optimization targets described.

As the health sector involves richer and more widespread data sets and has become more prominent and urgent, and more impactful these days, the partners decided to put the shipping pilot on hold and place emphasis on to smart medical and smart factory applicatin domains. Nevertheless, DESTINI will consider the possibility of collecting data for the shipping sector and will either proceed with setting up experimentations if time allows, or save the datasets for future processing in the context of the project's sustainability actions, e.g. being part of other proposals that may be submitted for funding in the future.

Another project that was introduced to DESTINI through its Dutch partners was VISOR (https://www.jads.nl/visor.html) which deals with safety in public spaces given the massive influx of people. VISOR attempts to tackle various issues, such as noise, crime, riots, harassment, pollution, environmental crimes, fires, undesirable social behavior, looting, public drunkenness, dehydration, drug incidents, bacterial infections, and a multitude of other dangers that affect the quality of life. The smarter collection and use of data in such cases, interpretation from the operation, as well as the more sophisticated collaboration based on data and interpretation, are according to VISOR potentially a decisive weapon in the fight against disorder and insecurity in public spaces. DESTINI will investigate whether the datasets collected or the techniques used in VISOR may also be utilized in pursuing the objectives of the SRIA over and above the primary application domains identified.

7 Summary

The purpose of this deliverable was to provide the research and innovation agenda of DESTINI. In this context it described in detail the research areas and the challenges that DESTINI plans to address and tackle.

The deliverable started with describing the general characteristics of Smart Data Processing and Systems of Deep Insight that are central to the SRIA and then provided some general concepts of DESTINI's Strategic Research & Innovation Agenda (SRIA). Then, a brief description of the SRIA and its main Joint Research Areas was provided, followed by a connection with two priority application domains, namely Smart Healthcare and Smart Manufacturing. Within these domains, the Strategic Research Areas that contribute to serving specific needs and challenges were discussed, describing in detail the plans, actions and objectives, while reference to additional priority research areas, such as IoT technologies and digital twins was also made.

The description of the methodology for application in the proposed domains was provided next, analyzing the strategic areas and the application development considerations in each of the domains of interest, followed by the plan to collaborate with strategic partners, stakeholders and industrial/market players.

The deliverable concluded by addressing some future plans for research and investigation of other hot topics and ongoing or finished projects with direct or indirect relationship to DESTINI.

This deliverable is linked to D5.1 - Strategy and action-plan for joint education, training and mobility, as it provides the scientific basis and the priority research axes which will guide the activities of the consortium towards establishing an efficient knowledge transfer scheme and will enable joint investigation and experimentation on significant problems in the area of Smart Data and Systems of Deep Insight.

8 References

Anderson, R.P., Lew, D. and Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological modelling, 162(3), pp.211-232.

Bardram, J.E., 2008. Pervasive healthcare as a scientific discipline. *Methods of information in medicine*, *47*(03), pp.178-185.

Calo, S.B., Touna, M., Verma, D.C. and Cullen, A., 2017, December. Edge computing architecture for applying AI to IoT. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 3012-3016). IEEE.

Dagli, C.H. ed., 2012. Artificial neural networks for intelligent manufacturing. Springer Science & Business Media.

Dimitrov, D.V., 2016. Medical internet of things and big data in healthcare. Healthcare informatics research, 22(3), pp.156-163.

Dubey, R. and Gunasekaran, A., 2015. Agile manufacturing: framework and its empirical validation. The International Journal of Advanced Manufacturing Technology, 76(9-12), pp.2147-2157.

Hutter, F., Kotthoff, L. and Vanschoren, J., 2019. Automated Machine Learning. Springer: New York, NY, USA.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H. and Wang, Y., 2017. Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology, 2(4), pp.230-243.

Lasi, H., Fettke, P., Kemper, H.G., Feld, T. and Hoffmann, M., 2014. Industry 4.0. Business & information systems engineering, 6(4), pp.239-242.

Lee, C.H. and Yoon, H.J., 2017. Medical big data: promise and challenges. Kidney research and clinical practice, 36(1), p.3.

Pickering, B.W., Litell, J.M., Herasevich, V. and Gajic, O., 2012. Clinical review: the hospital of the future-building intelligent environments to facilitate safe and effective acute care delivery. *Critical Care*, *16*(2), p.220.

Russom, P., 2011. Big data analytics. TDWI best practices report, fourth quarter, 19(4), pp.1-34.

Shi, W., Cao, J., Zhang, Q., Li, Y. and Xu, L., 2016. Edge computing: Vision and challenges. IEEE internet of things journal, 3(5), pp.637-646.

Sridhar, V., 2018. How MLOps Helps Automate ML Deployment Lifecycles. https://www.parallelm.com/mlops-helps-automate-ml-deployment-lifecycles/.

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H. and Sui, F., 2018. Digital twin-driven product design, manufacturing and service with big data. The International Journal of Advanced Manufacturing Technology, 94(9-12), pp.3563-3576.

Tao, F., Qi, Q., Liu, A. and Kusiak, A., 2018. Data-driven smart manufacturing. Journal of Manufacturing Systems, 48, pp.157-169.

Ullah, K., Shah, M.A. and Zhang, S., 2016, January. Effective ways to use Internet of Things in the field of medical and smart health care. In 2016 International Conference on Intelligent Systems Engineering (ICISE) (pp. 372-379). IEEE.

Yuhanna, N. and WaveTM, T.F., 2014. Market Overview: Big Data Integration. Forrester Research, Inc. Reproduction Prohibited.