DLT4PEOPLE

DR ILESH DATTANI & DR SHAUN TOPHAM

Personal data, which according to the EC definition is "*any information that relates to an identified or identifiable living individual* has been in the centre of attention for economic operators for ages, as the needs of retail and services are always coming back to the "know-your-customer" motto, where the preferences, desires, behaviours as well as other personal information is necessary to understand the target group and be in a position to offer the most efficient products and services, while at the same time save on investments which are doubtful of customer success. In fact, personal data and relevant information are key for offering novel services which are carrying the personalisation tag, going away from mainstream, mass production services and products which are lacking the real features that their target audience requires.

Recent innovations in sensors, activity tracking through wearable devices, Internet of Things (IoT) and Cyber- Physical Systems (CPS) technologies are building new valuable data streams that if tackled and combined properly - based on standardised procedures - could deliver new insights and services that could renovate sectors that rely on personal data. In the IoT and Web 2.0 era, there is an abundance of new data sources whose potential has not been yet fully leveraged; smart home sensors, wearable devices that enable activity tracking and safety psychological indicators, car sensors revealing driving patterns, user-generated content in social media like Facebook and Twitter, are only a part of a new rapidly evolving data ecosystem waiting to be explored.

Several studies have estimated the potential economic value of such data marketplaces at several billions of € annually in the EU alone. Coming to personal data, the Data Driven Marketing Institute reached the conclusion that with an estimated 3.5 billion active users, the total value of the market of personal data is approximately $210 billion, as advertising targeted using the personal data of an average user generates roughly $60.00. Looking back on companies which have been built on ideas of working with such data and are nowadays gigantic, it is evident that personal data includes much more value that one can imagine, as it can be used either on its own or in conjunction with data coming from other sources (individuals or not). However, this concept is inconceivable without enhanced movement of data as has been identified by the EU Single Market Strategy. In the last couple of years, there has been a stream of mishandling personal data from various perspectives which further hinders this flow.

Data ownership has been lost with the advent of social media platforms, which in their terms and conditions claimed ownership of whatever content the user stored in them. As such, content ownership and data provided by individuals, who were the rightful owners, has been shifted to such platform providers, who could use these data for their business needs and operation, without however offering any remuneration nor any other form of compensation back to individuals. Security has been compromised in many cases as malicious users were able to grasp personal data stored online, while communication channels for the exchange of

data, as well as the way data was managed and analysed did not employ strong security primitives in the sake of performance and cost reduction.

Furthermore, privacy has taken several blows, for similar reasons, such as in the case of services which seem to be provided for free but only work when the user consents to provide certain private information such as his current location. And although GDPR has been recently put into effect, a recent report claims there have already been over 95,000 complaints, and over 42,000 verified data breaches in Europe alone, resulting in the launch of 255 formal investigations already. These, with other recent negative development in this field and the emergence of "scandals" where personal data has been mishandled, such as the case of the Facebook-Cambridge Analytica scandal , which has led many individuals to not only take their privacy more seriously, but to go to the other extreme of not willing to share any kind of data, which hinders the development and the provision of personalised added value services.

All the above, resulted in a situation where the most critical of all challenges in data sharing is encountered, overarching all others, and that is a lack of trust. Most people believe that information is a valuable commodity but is of no use if we cannot trust the source or organize it in a meaningful way.

It is this lack of trusted and secure personal data platforms and the absence of privacy analytics methods which has decelerated the emergence of the division of the data economy which relies on personal data, as this unwillingness, and the complex methods and regulations that went into effect require large investments and a fundamental re- thinking on how such data should be managed, starting from the early stages of data collection, and going till the end of the value chain where insights are born and value is generated. Previously, technology did not provide a route for individuals or organisations below a certain size to become trusted contributors to information networks. However, recent developments are in position to handle back control over the use of personal data to individuals. And what is interesting to know is that these can give a huge boost to personal data sharing, as there is a general willingness of more than 70% of millennials (aged between 16-34) to share their data actual benefits, non-necessarily financial. To turn the tables and take of the breaks from the creation of such a data market that can benefit the whole of society, it is imperative to let the data owners (the individuals) decide how, how much and in which manner they would like to share it, following patterns and routes that can guarantee the privacy and security levels advertised, and of course allow individuals to retrieve a fair share of the value that their data generates, even when these data (or their derivatives) are used in secondary operations (e.g. in case an economic operator uses these data for further gains by re-selling it or re-uses the analyses). Regulations and technologies, which recently surfaced, pose as very good weapons and preventive mechanisms against mishandling such data and are in a position to re-gain the trust of individuals when it comes to data sharing, in order to ignite a new paradigm of personal data sharing which includes benefits for all the actors in the value chain, without compromising privacy and security. It is a matter of technology convergence, business innovation and cross- domain collaboration to address these challenges and utilise the existing and constantly surfacing structured and unstructured personal data to run novel data sharing functions and power secure and privacy preserving analytics that will renovate the existing ways of work of various industries, generate new knowledge and build an open ecosystem of constantly growing economy around personal data sharing and analysis.

As the EU paper on “An emerging offer of "personal information management services" report reveals, over the last 5 years a small number of initiatives have developed over the with the aim of “*providing platforms and/or services to individuals in view of handing back such control*.”. In general, there seems to be an agreement by people that personal data spaces should be promoted and be on an EU-wide level, to improve the transparency of data handling and ensure the compliance with the current and future EU data protection legal framework, building from ground an environment of trust and collaboration. Nevertheless, most of the initiatives and projects that are going on are rather small, and most are based on fragmented and domain-specific data which hinders the potential of building large data lakes of personal data, and thus there is no high interest from external stakeholders to access these data, nor of individuals to provide them. However, the overall potential of efficiently analysing even a small portion of the available data is immense.

Innovations in technologies on data management and analytics have been significant in the last couple of years, providing methods, tools and widespread infrastructure that allow to access, manage, preserve and analyse data at various locations (at source (e.g. edge) or centralised) and provide valuable insights for almost every domain one can think of. At the same time, major work has been carried out in the area of (cyber)security and privacy, deploying new infrastructure and methods that allow for stronger, more secure and thus highly trusted systems, while in the area of data sharing and marketplaces technology has reach the “plateau of productivity” of the Gartner hype-cycle, with numerous services focusing on IPR handling, unforgeability and in general value generation through data exchange. Addressing the concerns on privacy, ethics and IPR ownership over the DLT4PEOPLE value chain is one of the cornerstones of the project. Its goal to set, sustain and mobilize an ever-growing ecosystem for personal data and insights sharing and for enhanced collaboration between stakeholders relies exactly on DLT4PEOPLE personal data platform’s extra functionalities and methods for retaining data ownership, safeguarding security and privacy, notifying individuals of their risk exposure, as well as on securing value flow based on smart contract. Towards this direction, DLT4PEOPLE is committed to ensure strict compliance with relevant national and European legislations, such as the Regulation EU/2016/679 “General Data Protection Regulation”, the Directive EU/2016/1148 “NIS Directive”, and the upcoming Regulation on Privacy and Electronic Communications, “ePrivacy Regulation”, aiming at reinforcing trust and security in the Digital Single Market and at aligning with GDPR new rules, in the framework of a coherent modernization process of data protection and data privacy regime. The project will also adhere to H2020 ethical principles and recommendations of key EU-wide initiatives (such as BDVA), besides taking into account relevant policies, position papers and guidelines on emerging trends, such as data sovereignty. DLT4PEOPLE, will not only respect security, privacy, trust and data ownership policies and regulatory tools, but will be also engaged in the alignment and contribution to ethics-driven standardisation projects for prioritizing human well-being and fundamental rights in the data-driven economy.

DLT4PEOPLE will pursue its realisation through Seven SMART objectives divided into three (3) main categories:

- Business objectives (BO) that focus on the pre-commercial evaluation and validation of the solution, the expansion of the ecosystem, the spread of excellence gained and DLT4PEOPLE market entrance and sustainability.
- Technical objectives (TO) focusing on the delivery of the DLT4PEOPLE innovations, to be used as the basis for business exploitation and commercialisation through a concrete business plan.
- Scientific and Innovation objectives (SIO) focusing on the research to deliver a rigorous and self-standing methodology to drive the DLT4PEOPLE implementation and define its operational principles.

DLT4PEOPLE aspires to become one of the flagship personal data platforms which are fully compatible with GDPR and satisfy the privacy and trust consideration of users, with the competitive advantage as opposed to other approaches of employing a novel, fair and understandable value compensation mechanism that finally pays royalties to data owners. Being highly user driven and market oriented, it is expected to take advantage of a continuous period of simultaneously pilot operation and novel services delivery, to test and validate its offerings both for individual users as well as for organisations that seek personal data, and verify in a business context the value of them, towards rapidly entering the market after the project's finalization. The platform that will be delivered, will be deployed as centralised cloud-based infrastructure, accompanied by a Personal App (or core components that could be installed in other Apps the user already has), and this combination will accommodate on the one hand any individual and any personal data type the latter would like to store, manage and share, and on the other side various types of stakeholders coming from the industry or research, who seek to access, analyse and work on the data that can be shared with them. This approach automatically hints at a pan-European coverage, which could be easily expanded to become global, as personal data can be modelled on one unified data model that can be easily extended, and the infrastructure to be used can be accessed from any point and will be made available on various personal devices. The above points will be proven through the realisation of selected real-life use cases, driven by stakeholders of the value chain, and that currently seek and manage different kinds of personal data. In each one of the cases, specific data access and usage scenarios will be designed and developed, which will be facilitated by the DLT4PEOPLE App (or App core services) and Platform, taking advantage of both open, in-house and third-party big data and services, proving how the platform can assign and cash-out value in personal data and how collaboration and data value chains are in a position to multiply the wealth of already existing, but isolated sources. The validation of the solution over a period of 24 months will ensure the maximisation of the innovation potential and the knowledge transfer between the members of the consortium and the solution's refinement towards a go-to-market strategy.

Business Objective I: To deliver a pan-European Personal Data Platform, compliant with EU regulations, national laws and adopting high security and privacy primitives, which can be used by any individual for storing collecting and sharing, after consent, personal data (or derivatives), validated and populated through a set of representative, long lasting (24 months) demonstrators

Key Outputs: DLT4PEOPLE Infrastructure, DLT4PEOPLE cloud-based platform, DLT4PEOPLE Personal App, DLT4PEOPLE Demonstration Scenarios
Quantified Targets: 1 cloud-based Data Management and Analytics platform, 1 Personal Data Management and Analytics App, 1 Open source library of the Personal Data App components, 5 different demonstrators, 1 Business Validation and Impact Assessment Report, Impact Metrics as in Section 2.1
Means of Validation: Architecture and Design Blueprints, Technical Verification & Validation Plan & Results, Business Validation and User Acceptance Plan & Results, Feedback acquired through demos in events.

Business Objective II: To instantiate a novel business model for personal data and insights sharing where data is valued based on different modalities and is attributed to the rightful owners though trusted smart contracts that dictate usage and access rights, placing data owners as the starting point of the value chain and offering to them benefits and full control of their data and the way the latter is used

The core driving force behind the DLT4PEOPLE project is the need to provide control of personal data back to individuals and to attribute the data's (and its derivatives) value to the rightful owners, bringing upside down the current practice where data is generated by individuals but is owned by third parties, which in many cases results in loss of trust on how personal data is managed. Through this approach, individuals become key players in a data value chain where personal information plays a significant role in the business innovations, and the enterprises can open the door to new opportunities by being able to dive into a larger pool of data, that will help them to expand their business operation through new channels as the ones to be offered by the platform, as well as the generation of new services and products that directly spring out of the exploitation of the DLT4PEOPLE value chain. Therefore, all stakeholders to be engaged in DLT4PEOPLE can be characterised as "prosumers", as they can at the same time consume but also produce data and analytics. Those will be also able to benefit directly from these operations, either directly (through direct service and product offerings), or indirectly (through integrating these assets into derivatives of other products/service) and the overall data value will be attributed using fair schemes to the owners and the different operators, following a trusted smart contract mechanism based on distributed ledger technology. DLT4PEOPLE will bring forward a pioneering approach to realise such as multi-sided, omnidirectional value generation model, that will strictly and fairly define how value can be captured, produced, released and cashed

out, suggesting a new business for personal data management, transforming the way EU citizens and EU organisations manage and work with personal data, respecting both EU and national legislations on data and ethics.

Key Outputs: DLT4PEOPLE Trusted Data Management and Sharing Principles, DLT4PEOPLE Personal Data Sharing Business Model, DLT4PEOPLE Value Distribution Method, Smart Contract Patterns and Templates for Stakeholder Collaboration and SLAs
Quantified Targets: 1 inclusive analysis on how personal data is and should be managed, 1 PESTLE analysis of the DLT4PEOPLE ecosystem, 1 SWOT analysis of the DLT4PEOPLE platform, 3 different Business Model templated for personal data value sharing, 1 distributed ledger platform to facilitate transactions, 1 consortium wide exploitation plan, 1 inclusive financial strategy plan with cost breakdowns and future projections, 1 consortium wide sustainability plan, individual partners business models and exploitation plans
Means of Validation: In-depth state-of-the art analysis in personal data management and security/privacy/trust considerations. Business model and value sharing mechanism presentation. International journals' issues / conferences' proceedings. Cost structure analysis; Feasibility analysis; Breakeven analysis & risk management; Communication and collaboration for funding to ensure research continuity, market readiness and viability.

Business Objective III: To cultivate a trusted sustainable and ever-growing ecosystem where individuals can constantly enjoy benefits and organisations from diverse domains can take advantage of accessing (after consent) a multitude of personal data (or derivatives), renovating the way they operate by expanding their offered services.

The success of DLT4PEOPLE, as with any data platform, is tightly connected to the content it manages and its user base; two elements that are essential for the expansion and multiplication of the userbase and of the value stored and managed, as this is the only way that can guarantee that new users and new data in general is going to flow into the platform. The excellence of the tools and of the services, the perception of security and trust they can generate and, above all, the fair compensation mechanisms to be used is one side of the coin that can guarantee success and the interest of stakeholders; the other side is based on successful dissemination and marketing activities grounded on showcases (the demonstrators) that are be able to clearly demonstrate business impact and value at great ROI numbers. Therefore, DLT4PEOPLE has a clear, lean and agile dissemination and communications strategy to generate the necessary demand for project's results and accelerate the introduction of new stakeholders in the envisaged value chain, aiming at continuous expansion of both stakeholders, sectors and data as well. The project will generate such an open, trusted innovation ecosystem will work towards opening up both the

accumulated knowledge of the project, but also its infrastructure to new interested parties that would like to be engaged not only after but also during the project's implementation and demonstration phases, where external entities, apart from personal data in case these are individuals, could provide suggestions for improvements and features that will help to maximise the project's impact. In this context, specific attention will be paid to generate open source software and expose Open APIs, while community engagement of individuals, data scientists, developers, entrepreneurs and other stakeholders will be also pursued through the organisation of webinars and hands-on training workshops, that will allow them to experiment and develop their services to interoperate with the platform, or to integrate the Personal DLT4PEOPLE libraries to their existing Apps, and become the early adopters of the solution that will come outside from the consortium. Additionally, marketing and other dissemination and communication activities will be also intensively conducted, utilising online and offline channels, constructing a complete marketing kit eyeing the sustainability of the platform after the end of the project, attending relevant industrial (and academic) events to generate market demand, publishing articles in the press and scientific papers, liaise with other project to port technological offering and data, and develop an open web and social media strategy to reach a broader audience. To further enhance the visibility of the research results, the collected and existing research data will be subject to the H2020 Open Research Data Pilot, where applicable, for making the data FAIR: Findable, Accessible, Interoperable, and Reusable. A Data Management Plan will be prepared detailing data collection, handling, processing, sharing, curation, security, and used methodologies.

Key Outputs: DLT4PEOPLE Public Showcase and Web Presence, Marketing Kit, Exploitation and Marketing Plan, Dissemination and Stakeholders' Engagement Plan, Events and Workshops, Publications and Press material
Quantified Targets: Exploitation, Dissemination and Communication Targets and KPIs as in Section 2.2. Collaboration with >5 other projects and businesses handling personal data.
Means of Validation: Exploitation Reports, Sustainability Plan, Dissemination-Communication-Stakeholder Engagement Reports

Technical Objective I: To deliver an technical solution comprising of secure and trusted Data Management and Analytics cloud based platform as a Service, coupled by Personal Data Apps, for storing, manging, sharing and monetizing over personal data (derivatives) which can be used by any individual with the aim to capitalise on the real value of his personal data, without dropping control of ownership or loosing track of the usage methods, providing also constant awareness of the privacy, security and risks he may be exposed at by sharing such data.

DLT4PEOPLE will deliver an infrastructure which comprises of a cloud-based data management platform that is coupled by Personal Data Apps, acting on one side as a broker for personal data sharing for individual users who have their data on the Personal Data Apps and as an optional trusted cloud-based platform to securely store their data, and on the other side as a single point of access for data and analytics for third party organisations (industry, public sector, communities, etc) who are eager to search personal data catalogues, request access and eventually obtain such data by committing to specific, value bound smart contracts, either through manual actions, or through automated API calls that could link their systems to the DLT4PEOPLE cloud-based platform. The platform, as well as the Personal Data Apps will possess all the necessary functionalities to securely store, provide access and manage the whole lifecycle of the data, and will offer out-of-the-box analytics as a service, and useful visualisations over a modular and scalable architecture to allow interested parties to directly conduct analyses. The DLT4PEOPLE solution comes as a novel approach compared to existing solutions, as the latter suffer from isolation problems due to operating in silos, thus not allowing new knowledge to complement the existing results, as well as from other aspects such as data confidentiality, privacy protection and IPRs handling. All of the above are tackled within the DLT4PEOPLE concept, and due to the fact that the latter aims to be a broader personal data platform, extra functionalities and methods are employed to retain data ownership, safeguard security and privacy, identify risk and notify individuals of their risk exposure, secure value flow based on smart contract and adhere to the demands of national and EU-wide laws and regulations regarding personal data, taking in this way part of the burden that SMEs carry when they want to store and process such types of data. Moreover, DLT4PEOPLE takes a different approach to data analysis and access. Currently, data analytics and intelligence is offered in conventional and in most cases through strictly defined fixed queries, based on raw data, which on one side constraining the fantasy of end-users and their experimentation potential that could lead to new knowledge insights in unexplored dimensions, and on the other side possess a great threat to privacy and identification of individuals, demotivating the later to share their data. Overcoming these issues can be made possible through the roll-out of a platform that will rely on the adaptation of open source big data analytics algorithms to facilitate customised analysis based on user's input, data curation and semantic enrichment of data, leading into intelligence and insights that in turn will give birth to improved value assets used by the different stakeholders that utilise such data. And at the same time, the introduction of concepts such as digital twins of individuals, personas (e.g. fictional individuals that share the same, but obfuscated characteristics of specific groups of individuals), as well as analysis of second tier data (meta-analytics) are in a position to overcome the major concerns of people about their privacy. DLT4PEOPLE understands the significance of delivering such services to stakeholders in a timely, efficient and comprehensive manner, to allow the real value of data to be consumed properly and maximise the impact of the knowledge conveyed. DLT4PEOPLE will therefore deliver a personal data platform, that can be accessible via Open APIs to cope with the increasing requirements of data analysis third parties have, while supporting the inherited - by the integrated tools – privacy, security and value safeguarding. In this line, the Personal Data App to be developed will be also released as open source libraries, to allow interested third parties to integrate it to their existing applications, without thus disrupting their business models and ideas but allowing them to connect and benefit for the offerings of DLT4PEOPLE, and at the same time give the opportunity to their users to provide more data and capitalise on those. As a data platform, inclusion of new data and knowledge would be

possible following the DLT4PEOPLE semantic representation and data matching schemas, adhering also to the security, privacy and trust policies for personal data through the introduction of a secure data management framework that will allow data owners and data seekers to collaborate for the shared value generation and expansion of the overall solution. The infrastructure will support methods and tools for anonymization and privacy preservation, secure data channels, IPRs on data artefacts and data usage, to go beyond current practises in data sharing and handling, thus allowing stakeholders to securely exchange data. The overall data management and analysis will follow the latest containerisation principles, to safeguard critical data, sharing only the agreed elements based on the smart contracts and the accompanying service level agreements to be signed by the interested parties.

Key Outputs: DLT4PEOPLE cloud-based platform, DLT4PEOPLE Personal App, DLT4PEOPLE Personal App libraries, Data Model, DLT4PEOPLE Open API, Data brokerage engine, Documentation and Usage manuals
Quantified Targets: 1 SaaS Platform, 1 Mobile App, 1 library for Mobile App development, 1 expandable knowledge representation (in the form of a knowledge graph or ontology) for the personal data, 1 brokerage platform compromising of 2 separate layers, 1 secure cloud-based storage facility, 1 cloud-based data analytics engine, 1 access control engine, 1 risk management service, 1 visualization library, 1 personal data catalogue, 1 business brokerage engine
Means of Validation: Architecture and Design Blueprints. Technical Verification & Validation Plan & Results. User Acceptance Plan & Results. Feedback acquired through demos in events.

Technical Objective II: To integrate existing approaches, tools, libraries and components that allow handling of personal data in the way they should be preserved, accessed, valued, and controllably shared, guaranteeing high quality results which can support rapid prototyping, traction generation, fast market entry and sustainability

DLT4PEOPLE will work on creating specific services that are relying on the latest development of secure and trusted data storage and sharing technologies, as well in data analytics, integration points and APIs on top of existing open source tools and computational infrastructures of high TRLs (delivered through operation tools and previous research attempts) to defining a state-of-the-art solution that is able to satisfy all needs for trusted and fiar data sharing and infuse the necessary intelligence to existing analyses by blending different data streams and knowledge together. Data technologies are at the moment witnessing an incredible bloom, with new tools coming out every day in the different sectors, ranging from new sensors, data communication channels and engines that process all the collected information. DLT4PEOPLE will closely monitor all these developments and will select

for integration and consideration the crème-de-la-crème artefacts of the above developments, that are highly relevant to the needs of the personal data sharing and create the respective bundles that will be used to drive the operation of the system. DLT4PEOPLE will provide a data storage, management, analysis and integration infrastructure to address a) the identification, access and interoperability among sources and sensors which output personal data built by different manufacturers, on a variety of technologies and targeted for a plethora of applications, b) the application of semantic technologies coming from well-established projects in the domains of Linked Data, to support schema matching of such diverse data streams, both of structured and unstructured nature, retrieved from the abundance of sensors and personal data streams available, thus tackling both real-time data and batch data stores, c) the trustworthy and secure sharing of data which entails increased complexity regarding access control and risk management, focusing also on the provision of risk awareness indicators to individuals regarding their shared data, d) the fair and trusted documentation of transactions and IPR management between data owners and data seeking organisation, that can result in monetary or other types of compensation, e) the analysis of data that derives out of interlinked and semantically enriched personal data at multiple detail and security levels, as this requires novel approaches to data segmentation, sampling, flattening and processing to achieve the required performance criteria

Key Outputs: Modularised Services and Tools for data management and sharing as part of the platform, a unified data management service to interconnect all other components, improvement and integration of technical data infrastructure solutions supporting both secure and trusted data exchange and retention, a novel paradigm for the documentation and IPR handling of conducted exchanges
Quantified Targets: Run 200 questionnaires for needs elicitation, 1 platform architecture, designs of all software bundles, integration of at least 8 TRL>7 technologies, support for cutting edge technology for security and trust by design, support for modern analytics algorithms on plain, multiplexed and encrypted data
Means of Validation: DLT4PEOPLE Methodology, DLT4PEOPLE MVP, Architecture and Bundles Design, DLT4PEOPLE platform releases, Evaluation in the Demonstrators

Scientific and Innovations Objective I: To deliver an innovative, secure, privacy preserving, IPR respecting, and fair compensation data exchange methodology, propelling the creation of a joint venture of personal data owners and data seeking organisations

The provision of the DLT4PEOPLE to both individuals and to data seeking organisations will establish a common ground for collaboration between stakeholders, leading to partnerships, allowing data owners to finally get compensated for the assets they provide and business entities to further advance their operations by blending internal intelligence with external information pieces (those of personal data) which would otherwise not be accessible. The overall approach is built in a way that it does not compromise privacy and security over data analysis, mostly due to the establishment of proper data usage policies and novel compensation schemes that will be put in place and will be based on algorithms that will

calculate a valuation the type of data shared the granularity of it (raw data, anonymised data, obfuscated data, analytics extracts), as well as other factors that accompany data. These features are expected to open up new opportunities for knowledge harvesting, collaboration and also for building new profit centres by externalising and commercialising personal data and information, and for supporting the business objectives of the project, resulting in an ever-growing active community of prosumers.

In this context, DLT4PEOPLE will deliver a conceptualisation framework based on fair, secure and trusted data (and its derivatives) sharing principles, powered by blockchain technology innovations that will facilitate knowledge exchange, setting in motion the appropriate security, privacy, IPR policies to resolve on-the fly how data can be handled by each stakeholder group, based on its content, its value and peer-to-peer agreements that will be reached between the collaborating entities. This will foster the creation of a trustful and rigorous data sharing community, where value is shared and exchanges amongst its members following a fair distribution of resources and costs.

Key Outputs: Value Chain definition, Personal Datasets and Data sources, semantic representation of Personal Data, update on existing semantic vocabularies and contributions to LOD, data analysis algorithms
Quantified Targets: Inclusion of personal datasets of 5 different sites/demonstrators, 20 types of personal data categories supported, Compatibility with at least 20 types of data sources (sensors, IoT, APIs, wearables, records, etc), 12 known analytics algorithms supported, 10 existing vocabulary standards reused

Means of Validation: DLT4PEOPLE Methodology, Design Documents, Data Value Chain Definition, Demonstrator Applications and Evaluation .

Scientific and Innovation Objective II: To successfully link novel trusted and security-by-design data mining, management, analysis and sharing techniques, with legislation- and ethics-driven functions, facilitating both privacy and trust preservation, risk situational awareness, easy access to, and usage of valuable information and fair compensation models for all the actors of the value chain.

One of the key challenges when it comes to handle personal data is to find the equilibrium point where privacy and security is not sacrificed at the sake of performance and analysis, without at the same time hindering the performance of the latter as well. Currently, many systems that handle data are being refactored, due to stricter data management directives and laws (such as GDPR), however as most of the actions are categorised as ‘patches” they continue suffering from vulnerabilities that have to be constantly tackled and mitigated, without offering a final and viable solution in the long term. DLT4PEOPLE comes as an

infrastructure that offers privacy and security by design, building on the top of the research conducted by the consortium partners in the areas of security, privacy and application of regulatory frameworks for those issues, and is effectively combining the needs of individuals regarding their data and those of organisations that are eager to access and manage those. As such, barriers that have to do with those domains can be overcome, while new innovations such as searchable encryption analytics pave the way for placing a stronger guarantee seal on the final output regarding the efficiency and the effectiveness of the solution, also from the side of business value, performance and intelligence delivery. At the same time, the overall concept is empowered by a new methodology for value generation and remuneration, which relies on innovations that have been witnessed in domains such as fintech in the last years, and should be ported to the data management domain to accelerate the growth of the domain and motivate all stakeholders to get engaged. DLT4PEOPLE will deliver a framework that spans from the semantic enrichment and interlinking of personal data towards generating unified and reusable data artefacts, available at various security levels to the different value chain actors, to the last stages of data analysis. These activities will be based on the definition of a unified personal data schema, extending existing standards and ontologies, and will take into consideration conventions on how data, even in when encrypted, can be searched to denote business opportunities , how data or its derivatives can be shared and delivered to interested parties in a secure and trusted manner adopting TPM technology, following a series of actions that specify the degree of privacy and security to be supported, the privacy risk exposure calculated based on previous knowledge depending on the data already available and shared, the retention of data ownership and the usage rights that will be directly linked with compensation schemes which will go into immediate effect. The outputs of those scientific actions will be transformed into technical solutions (see TO.II above) which will allow the exchange and documentation of data in a unanimously understandable manner.

Project Concept – Overall Approach

DLT4PEOPLE aims to deliver a novel framework and architecture that leverages personal data, coming from diverse sources (sensors, IoT, wearables, data APIs, historical data, social network data, activity trackers, health records, demographic profiles, etc.) to help individuals construct their unified personal data hub, collect at a single point all of their personal data in a secure and trusted manner, and retain ownership and control on what to share and with whom, receiving also compensation for the artefacts they place at the disposal of other third parties. In turn, third party organisations (companies, public sector, NGOs, etc.) arrive at a position where they can request and get access to tons of personal data, which can complement the ones they already manage and that can be used for generating more efficient, effective and value added services, engaging with individuals into an entirely new way for data sharing, which generates trust and an increased feeling of collaboration, as the data owners (e.g. individuals) become the centre of attention and the most important partner and collaborator of those third parties.

As a result, DLT4PEOPLE aspires to provide controlled access to one the largest gold mine of data (that of personal data) and apart from facilitating the need for establishing a prominent

personal data platform, lays the bridges to link with industrial data platforms, as the access and data usage policies to be selected could lead the way to allow third party organisation to capitalise on the collected data or on the analysis outputs, and through this process still allow value and compensation to flow back and ultimately reach the data owners, upon the data of whose such new services and new business innovations have been built. Moreover, as currently various services and application rely on personal data, but are focusing only on a small fraction of the whole domain (for example smart phone Apps that store activity data), the approach taken by DLT4PEOPLE has been carefully selected to respect the business models of such enterprises and their need to retain such data through their own channels, and allows them to connect to the platform through the integration of specific libraries, that are non-intrusive to the already established way of operation, but can allow for broader data capturing and sharing.

The data value chain that is tackled by DLT4PEOPLE can be seen as a structure of a trusted cycle, which includes:

- **Primary Personal Data Providers (Individuals)**. This tier includes all the individuals which are generating and collecting their personal data from various services, devices and applications. It is these data which is considered

  "personal" and constitutes the core data of that is of interest of the DLT4PEOPLE project.

- **Economic Operators.** These are data seekers (also titled as 1st-tier economic operators), that look for enjoying business intelligence based on Primary Personal Data. In this tier, data seekers (organisations of any type) are able to work on the data of the first tier (primary data) and combine them with other types of data they have to create new datasets or relevant derivatives (insights, reports, etc)

Apart from these two main actors, DLT4PEOPLE aims also to attract a third category, which can be seen as **2nd-tier economic operators** that provide data and services based on analytics or data that is shared and generated by the economic operators which belong in the core data sharing cycle (1st-tier). Such stakeholders are interested to provide services that are based on data reused/resold/etc. In a typical scenario, the value generated by organisation of this tier does not flow back to the data owners, as business deals are restricted between the Tier 2 (demand) and Tier 2 (supply) entities. Exploring the value that can be generated through these tiers, DLT4PEOPLE aims to capitalise on such existing modern technological breakthroughs in the areas of the data-driven economy, and roll out a completely new, trusted and secure value chain of interrelated data streams coming from individuals (personal data) in order to revolutionise the way personal data can be managed and shared with interested stakeholders, demonstrating a huge, but realistic economic and societal impact that can being achieved by introducing a fair and well-balanced economy build around personal data, that will be linked to various sectors whose services rely on, or are interested in sourcing and analysing such data. As such, main outputs of the project would be novel GDPR compliant services that allow individuals to capture, access and manage their own personal data and decide the way they would like to share them (including all relevant aspects

that are present on any transaction), which for example could be responding to certain requests that may come from 1st or higher tier data consumer organisations, publicly displaying their availability or simply providing their data openly for demonstration and experimentation purposes.

DLT4PEOPLE builds on the premise that innovations in sensor technologies and in APIs are allowing for new, fast- paced data streams even when these concern personal data. If tackled properly and combined based on standardised procedures, these data could deliver new insights and services that could renovate a large set of domains that are directly or indirectly linked to services or consumer for the public. As depicted in the previous table, the core data sources of the DLT4PEOPLE project fall within the scope of the 4 V's of big data, Volume, Variety, Velocity, Veracity. This approach will be based on the introduction of Personal DLT4PEOPLE, as Apps residing at the users side (at the edge), on smartphones, set-top boxes, or other infrastructures like Raspberry Pis, Arduinos, etc.), which will include a secure storage facility to hold the data and added value services to search them, assess them, and share them in a trusted and secure manner, following recommendations on risk exposure metrics and the creation the of smart contracts between the data owners and the data seekers (organisations). As such, the by-design secure architecture of such DLT4PEOPLE solution ensures that information from various sources is retrieved and securely stored at the edge (user's device), following modern encryption techniques to guarantee privacy and total information security. Personal Data Vaults are also coupled with a small-scale data insight engine, which resides on the edge and are able to conduct in-situ analytics, in case a user decides to share analyses of their data, rather than the raw data (see Table 1-2 below).

Table 1-1. Types of Data and Properties

Personal Data Type	Data Sources	Format				
Demographics	ID Documents, Accessible Public Records, User Entries	Structured/Unstructured/ Sensory				X
Physical Activity Data	Wearables, Activity Trackers, Applications and APIs (FitnessPal, etc)	Structured/Sensory	X		X	X
Health Data	Health records, Wearables, Applications and APIs (Andaman7, Apple Health Kit, etc)	Structured/Unstructured/Sensory	X		X	X
Location Data	Smartphone Location, Applications and APIs (Foursquare)	Structured/Sensory	X		X	X
Service Usage Data	Applications and APIs	Structured/Unstructured/Sensory	X	X	X	X
Smart Home Data	Energy Dashboards, Sensor Data					

		Structured/Sensory	X		X	X
Social Activity Data	Twitter, Facebook, LinkedIn, Instagram	Structured/Unstructured	X	X	X	
Sentiment Data	Twitter, Facebook, Instagram	Structured/Unstructured	X	X	X	
Travel Data	Travelling Applications	Structured/Unstructured/Sensory				
Financial Data	Tax Registries, Bank Feeds, Online Payment methods	Structured			X	X
User Generated Content	Photos, Videos, Etc.	Structured/Unstructured/Sensory	X	X	X	X

In addition, these data, may be optionally (at users will) fully or partially replicated in a secure, trusted and encrypted manner also in the cloud-based DLT4PEOPLE centralised platform to expose certain metadata that can be accessible by third parties (either directly or through searchable encryption analytics) and facilitate the exchange and analysis of heavier and more performant intense analytics. These actions are made possible by the employment of secure communication channels, novelties in encryption and analytics domains (searchable encryption analytics) and methods that guarantee the anonymity preservation such as the DAA protocol which will be offered to the different entities taking part in the platform (e.g. at the edge nodes but also in the central repositories) through hardware/virtual/software TPM implementations. At the same time, anonymised and obfuscate unencrypted data can be also uploaded via the same technique to the cloud infrastructure of DLT4PEOPLE, resulting in the generation of a digital twin "persona" of a user, which possess same characteristics but cannot be traced back to the actual individual. In this scenario, data under DLT4PEOPLE can be shared and made available to third parties in a various number of combinations of their modalities, some of which are the following:

- Data Location (Edge/Cloud) - Storing data on the users' side, or on the cloud engine to allow heavier analytics methods to be applied, or for more data to be discoverable and available through operations performed by the DLT4PEOPLE Cloud Engine.
- Data Sharing Amount (Full/Partial) - Sharing all, or certain parts/types of the metadata and the data, incorporating other parameters such as location, timeframe.
- Data Encryption (None/ Searchable Encryption Protocol/Fully Encrypted) - Encrypting the data for maximum privacy, allowing searchable encryption methods to be applied on the cloud-based platform, or keeping data unencrypted in the platform.
- Data Anonymisation (None/ Digital Twins/Anonymised for "Personas"/) - Sharing the actual data of individuals, creating anonymised "digital twins" of individuals through

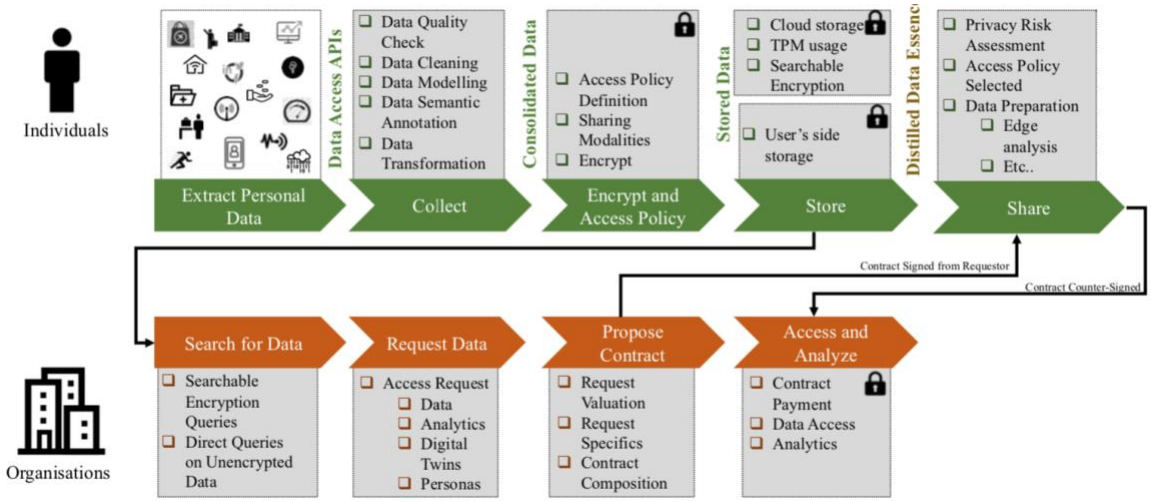
anonymisation and data obfuscation mechanisms, or sharing anonymous data to be grouped under “personas” which contain data from similar individuals.

DLT4PEOPLE takes also a different approach in sharing by allowing individuals to share:

- Data, based on the properties above, which do offer to economic operators more accurate results but are resource greedy and in many cases include a higher degree of privacy risks
- Data Analytics, which can be performed either on the edge or on the cloud-based DLT4PEOPLE platform, which may provide the same accuracy as sharing plain data, but improve the privacy level of individuals
- Complex Insights, which are combination of analytics, that further improve privacy but provide valuable information to interested stakeholders.

Every such data “publishing” transaction between individuals and the DLT4PEOPLE cloud-based platform is logged in a distributed ledger, called the Private Ledger, which clearly identifies through hashes data that belongs to users, without however disclosing the actual identity of each user to external entities. Moreover, to ensure privacy also from the side of the platform, the Direct Anonymous Attestation is utilised, to deem data senders anonymous. As such, the DLT4PEOPLE cloud-based platform can hold a registry of all available datasets, and also perform operations that group together diverse datasets putting into display a data catalogue that is searchable by other entities. In turn, other organisations are able to search for data (through searchable encryption protocols or through typical queries on non-encrypted data), and even directly request the platform to provide them with specific datasets, grouped datasets as well as insights, by constructing complex queries. All such data access requests are logged in the DLT4PEOPLE Open Ledger, which is responsible for creating smart contracts between such organisations and the DLT4PEOPLE cloud platform, as the latter functions as a proxy operating on behalf the individual users. Once the contact is created, individuals are presented with the contract terms which amongst other refer to the type of data (or derivatives), the modalities requested (e.g. aspects such as the price, the usage terms, the retention period, etc. of the data to be shared. In parallel, a risk assessment and privacy exposure report incorporating understandable privacy metrics is served to the user, allowing him to assess whether the data to be shared, in conjunction with his previous transactions and with other external factors, may jeopardise its privacy or not. Once these contracts are validated, the Access Policy Engine is used to pick point and serve the requestors with access to the information they need, which can be accessed through the platform’s analytics engine. At the same time, the contract between the platform and the data requestor is cashed out, and another contract between the platform and the individual is automatically instantiated in the Private Ledger, in order to transfer the value from the platform’s wallet to the wallet of the individual user.

In detail, as depicted in the concept figure, DLT4PEOPLE will enable an iterative process to be undertaken by different stakeholders who wish to leverage the power of the personal data value chain offerings:



- Data Providers, which in the DLT4PEOPLE case are individuals, are able to extract, collect in one place and securely store their personal data, and take control of their usage. These data can be published in various modalities to the DLT4PEOPLE platform and placed on display so that interested organisation can find them, request access to them and get shared based on trusted contracts with well-defined terms and conditions.
- Data Requestors (e.g. economic operators/stakeholders such as businesses, public sector organisations, etc) may explore extracts or metadata of the data that can be made available by the individuals, request access to them and once granted experiment in an analytics sandbox within the DLT4PEOPLE platform, by combining the extracts of data with their own (private) data and by running various analytics algorithms.

Based on the above, the services to be offered by the project include:

- Holistic personal data management services, including collection, mining processing, normalization, formatting and availability at individuals' personal devices level as well as on secure data vaults on the cloud
- Smart interlinking of personal data to open, linked as well as proprietary data following Linked Data principles and openly (re-)publishing non-sensitive and business critical information to the LOD community
- Novel data security and cryptography, data anonymisation and privacy preservation, remote attestation and trusted data exchange through the utilisation of TPM technologies between the Personal DLT4PEOPLE and the DLT4PEOPLE cloud-based engine
- Privacy risk assessment methods that offer a "situational awareness" picture to individuals with easy to understand privacy metrics, revealing the true risk exposure factor of individuals based on the shared data
- Privacy preserving and data security retention mechanisms, to accommodate the generation of anonymised "digital twins" of individuals, as well as specimen clusters ("persona groups") powering group analytics that contain valuable insights without violating privacy principles

- A twin fold data brokerage engine to cater for IPR and data license safeguarding, documenting transactions in a privacy preserving, yet undisputable and unforgeable manner, facilitating compensations schemes with third parties (that support the shift to future monetisation streams) through the instantiation of multi-layer real-time micro-contracts specifically tailored to the needs of data sharing, redistribution and utilisation, constructing a bridge between personal data and industrial data platforms.
- Smart balancing of analytics methods to accommodate Edge Analytics as well as centralised operations depending on the degree of data volume, velocity and variety, always in conjunction with the security and privacy modalities allowed by the individual for each kind of analysis
- Provision of intuitive analytics, reports, smart dashboards and visualizations tailored to the needs of each stakeholder of the domain, including the individual, as well as generic ones for wider use by any interested organisation and by the public

In summary, DLT4PEOPLE will capitalise on the latest advancement of data management and analysis methods, as well as on breakthroughs in the areas security, privacy and trust mechanism, with a clear target to flourish in the Digital economy and deliver an personal data management infrastructure able to transform the way personal data is treated in Europe today, and demonstrate huge and realistic economic and societal impacts for all engaged parties.