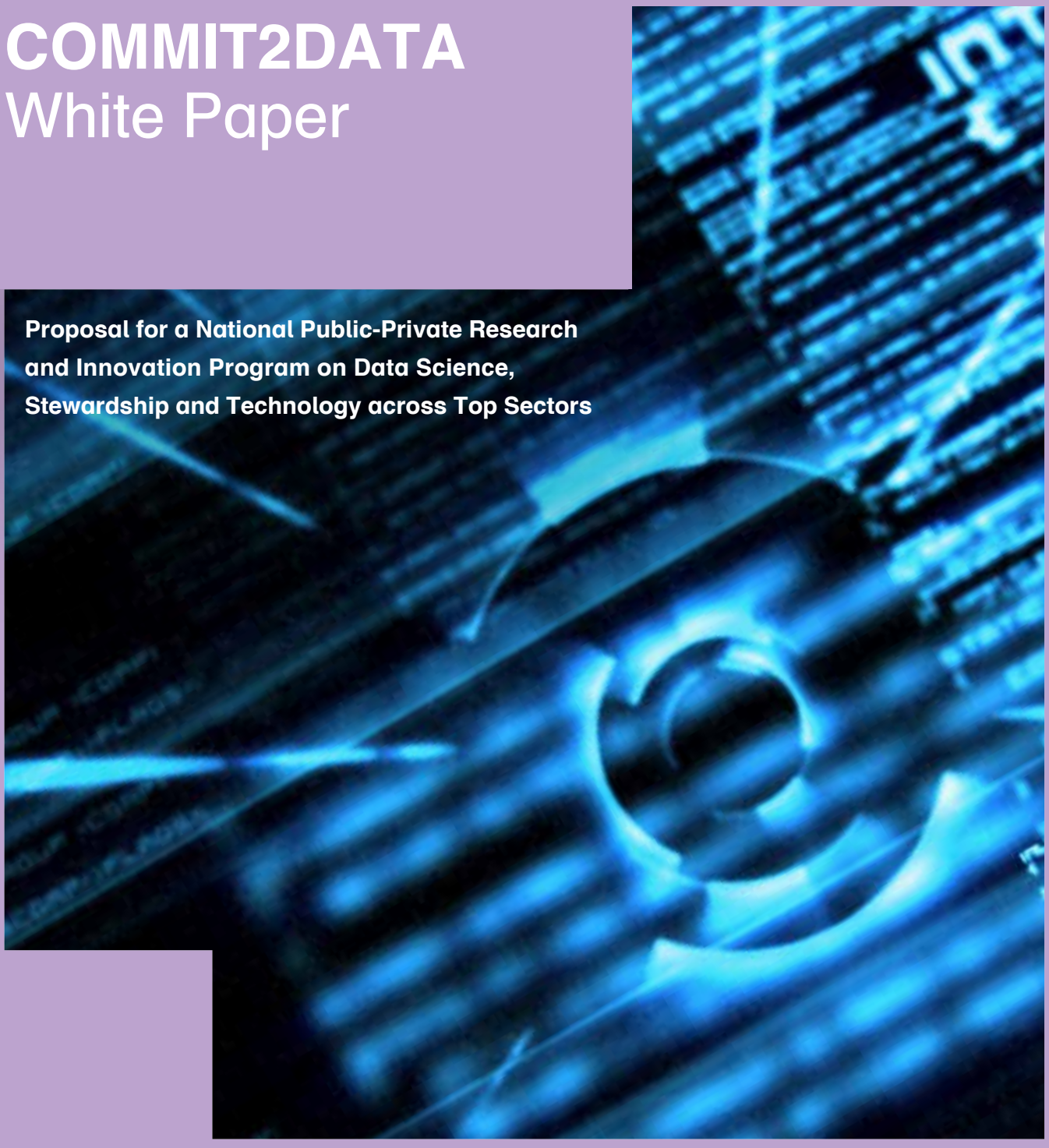
dutch
digital
delta

COMMIT2DATA
White Paper



Proposal for a National Public-Private Research and Innovation Program on Data Science, Stewardship and Technology across Top Sectors

connect and create

dutch
digital
delta

Team ICT

René Penning de Vries, Captain
Inald Lagendijk, Captain of Science, Delft University of Technology
Ineke Dezentjé Hamming-Bluemink, FME-CWM
Ben Woldring, Bencom Group
Gerben Edelijn, Thales Netherlands
Mark Bressers, Ministry of Economic Affairs

Concept & design

Dune Reclamebureau - www.dune.nl

Photography

Wavebreakmedia/Shutterstock: "*Program code*" (cover),
"*Holding hand out in presentation against data center*" (page 13)
Kubais/Shutterstock: "*Close up of hard disk with abstract reflection*" (page 7)
Zadorozhnyi Viktor/Shutterstock: "*Close up computer motherboard*" (page 27)
Bikeriderlondon/Shutterstock: "*Climber using laptop on mountain peak*" (page 37)

September 2015

http://www.dutchdigitaldelta.nl

Table of Contents

Executive Summary

Big data is the 21st century “natural” resource for innovations in society, economy and science. Pervasive technologies such as Internet of Things, high tech sensors, digitization of collections, and social media lead to massive amounts of collected data. The value ascribed to these big data is found in unexpected patterns hidden in the data, and the causal relations, predictive powers, and decision support that can subsequently be attained. Unearthing the value of raw big data as so to obtain hindsight, insight, and foresight in any application domain requires breakthroughs in advanced ICT science and technologies including machine learning, data mining, visualization, semantics, data bases, software engineering, and data protection.

The COMMIT2DATA proposal for a national public-private research and innovation program on data science, stewardship and technology across top sectors has been developed by Team ICT as part of the Knowledge and Innovation Agenda ICT 2016-2019. The proposal has been developed in close collaboration with stakeholders and the various Knowledge and Innovation Agendas in urgent economic and societal challenges that are susceptible to ICT innovations. COMMIT2DATA aims to maintain and strengthen the Dutch top-5 knowledge position in big data. The program bundles and focuses data science expertise, resources and funding in private, public and government sectors. The program leads to a durable contribution to the human capital agenda in the field of big data, it will further the results of prior investments in ICT and big data, and it will contribute to regional anchor points for valorization and dissemination of big data.

The program builds on the excellent Dutch ICT knowledge basis, consolidated data science research and valorization efforts at universities and other knowledge institutions, and proven successful public-private collaboration in ICT such as in the COMMIT precursor program. COMMIT2DATA will stimulate high-tech entrepreneurship and will propel the Netherlands forward to becoming a big data main port.

COMMIT2DATA is structured as three integrated program lines. The first program line entails pre-competitive use-inspired research in the field of data science, stewardship and technology. In this program line research challenges common to top sectors are addressed such as the semantics of data, computational complexity, data protection and human information overload. Use-inspiration for the data science, stewardship and technology research mainly comes from “data for life”, “data for energy transition”, “data for smart industry” and “data for security”. Research projects that will eventually be formed within the COMMIT2DATA program will cover a specific data science, stewardship and technology challenge across multiple sectors as so to maximize lateral knowledge transfer. The second program line covers valorization sprints. It provides the main vehicle for strengthening big data knowledge and technology in companies and startups across all top sectors. Finally, the third program line, dissemination, focuses on disclosing big data knowledge, tools and solution to SMEs by instruments such as “data factories” and “big helps small & small helps big”.

The envisioned budget of the COMMIT2DATA program – *excluding matching first money stream budget of knowledge organizations* – is € 154 million for a period of 5 years, to be funded from private companies, and regional, national (NWO, ministries) and European (H2020) programs. Overall the program aims at 30% private funding.

Proposal

Proposal for a National Public-Private Research and Innovation Program on Data Science, Stewardship and Technology across Top Sectors

Chapter 1

Ambition, Context, Structure

Ambition, Context, Structure

Big data is widely regarded as the next frontier for innovation, competition and productivity. It is seen as the 21st century's "natural" resource for innovative services and processes with enormous added value for society, economy and science. Big data is instrumental in successfully making the transition to innovative solutions wherein the (consumer or business) end user is central. For example when looking to the areas of personalized care, energy management, and highly personalized products. The Knowledge and Innovation Agenda ICT (ICT Roadmap) 2016 makes evident that both large and smaller businesses and organizations in nearly all Dutch economic top sectors see extensive benefits from breakthroughs in the field of big data, and they are therefore moving to invest in the Knowledge and Innovation Agenda ICT 2016. One important big data breakthrough that is consistently mentioned is the unearthing of patterns in the data as so to acquire hindsight, insight and foresight in a particular application domain. Another notable breakthrough is the storage, interconnection and protection of big data across multiple sources, users, companies, and even geographic locations.

These breakthroughs will lead among others to better business intelligence, more informed decision support, better use of scarce resources, a higher degree of product personalization and automation, better understanding of customer behavior, and a chance to gain deeper insight in root causes of diseases, production defects, and security threat. At a broader scale, breakthroughs contribute to progress on sustainability, (bio-) diversity, and citizens' and society's vitality. New companies will emerge that create value by addressing generic big data challenges as well as sector-specific applications. Achieving these big data breakthroughs successfully requires a solid knowledge, valorization and dissemination basis in the field of data science, stewardship and technology, or "data science" for short. The COMMIT2DATA national public-private program (PPP) program aims to bring together knowledge institutes, government and companies within a number of strong economic sectors. The aim herein is to jointly maintain and strengthen the Dutch top-5 knowledge position in data science. The economic and societal sectors that COMMIT2DATA builds upon have been selected based on data science urgency expressed in the Knowledge and Innovation Agenda ICT 2016, and their susceptibility to innovations with information and communication technology (ICT) sciences. The collaborative and coherent effort across top sectors drives focused data science research, maximally leverages use-inspired research results, maximizes the valorization opportunities for research results in different sectors, and ensures the optimal use of private and public research investments.

Two Dimensions of COMMIT2DATA

The structure of the COMMIT2DATA program is shaped by the observation that any big data public-private program must bring together (1) excellent data science knowledge with (2) specific application domain inspiration and knowledge. This is essential because the field of big data is delineated in two dimensions.

The first dimension considers data properties and objectives. If the focus is on the volume, **velocity**, and **heterogeneity** of the data, the challenge is to find algorithms for efficient, massively parallel, robust and cognitive processing that disclose the deeper meaning of the data. If the **quality** of the data is a dominant issue because (uncontrolled) data is exploited that is external to the service *– such as self-reported or social media data –*

the challenge is to establish the reliability and trustworthiness of the conclusions arrived at. If **theft**, **privacy**, **access** and **longevity** of sensitive data are the main considerations, then data science focuses on data management and protection. And finally, the **impact** of the data can be an economic value or intellectual insight. Each of these data aspects comes with its own body of knowledge and scientific challenges. Together they define the science of data, stewardship and technology in any application using big data.

The second dimension is the **application** domain and **context** in which big data is used. Big data does not exist in isolation, there is always a context in which the data is generated. In order to successfully reap the value of big data – *be it economic value or intellectual insight* – some degree of contextual information is needed; one needs to understand the specific properties and limitations of the data and its intended use. Applications may be very different because of the dissimilarities of the sectors themselves. But the underlying subset of data science challenges is often very similar indeed. Hence, COMMIT2DATA aims at a maximally effective program by seeing application needs and development as well as the data science research challenges from the perspective of commonalities in the properties of the big data.

Big Data Position of The Netherlands

The Netherlands has a strong position in the field of ICT and ICT sciences which can be leveraged to bootstrap the COMMIT2DATA national program. This position is due, on one hand, to the affinity of the Dutch with advanced ICT technology and applications, and to the high penetration degree of world-class ICT infrastructure, including SURF's higher education and research ICT infrastructure, and the Amsterdam Internet Exchange AMS-IX, a backbone of the global internet. Big data research and applications use this infrastructure in valorization and in creating economic value. On the other hand, the Netherlands has an internationally strong knowledge and practice opportunity in big data for creating intellectual value, with valorization potential. ICT sciences in the Netherlands are at a very high level. Especially in data science a substantial number of individual national (VIDI-VICI) and European (ERC) research awards have been won, and several research teams are counted among the world leaders. Academic expertise centers have been established at the universities in – *amongst others* – Eindhoven, Amsterdam, Delft, Groningen and Leiden. Furthermore, Dutch data science talent drives innovation. After graduation, PhD students in data science frequently found startups and join world-leading innovators such as Google, Twitter, Facebook and Microsoft[1].

The Netherlands has a culture of collaboration among sectors and in public-private partnerships, as witnessed, for instance, by the COMMIT use-inspired research program. This € 110 million public-private program has attracted over 30 additional private partners since mid-2014 in addition to the initial number of 50 private and 20 public partners. Over its course of running, COMMIT has evolved from a generic ICT use-inspired research program to a precursor of a data science program. In international context Dutch data science specialists collaborate with IBM in the initiatives CHAT and ERCET covering a broad technological spectrum ranging from astronomy and humanities to life sciences and energy.

Several initiatives have emerged since 2014 that emphasize valorization and dissemination of big data techniques in small and medium sized

[1] For instance, Dr. Mishne – *graduated from UvA* – was Director of Search at Twitter (2012-2015).

enterprises (SMEs), such as TNO's Almere Big Data Value Center and the "ICT doorbraakproject big data". SMEs and start-ups are flexible in their strategy, and have the potential to pivot to new technology and build new business on data science results. Especially in ICT and big data, large companies innovate by absorbing fast moving start-ups and SMEs. At the same time it is important to note, that a 2014 study reported upon by Harvard Business Review shows that the position of the Netherlands in ICT innovation is threatened by a slowdown in private investments in innovative electronic services, skills, new markets, and start-ups. Investment in the COMMIT2DATA program by knowledge partners, government and businesses will intensify the Dutch data science ecosystem and propel the Netherlands forward towards becoming a "Big Data Mainport".

The European Commission and several countries surrounding the Netherlands are moving forward rapidly by investing in PPP collaborations on big data. Starting from 2014, governments have invested from € 75 million (Ireland, Austria) to € 250 million (U.K.) in big-data programs, and the EC has recently announced a € 500 million Big Data public-private partnership in collaboration with the EU Big Data Value Association. All these public investments aim at building up knowledge, driving valorization, and widening dissemination in data science. The goal is to accelerate gains in value across a range of economic sectors.

Ambition of COMMIT2DATA

COMMIT2DATA directly connects to the Knowledge and Innovation agenda ICT (ICT Roadmap) 2016 and data science challenges submitted to the National Research Agenda (NWA) in 2015. The program emphasizes use-inspiration for data science research from major economic sectors and societal challenges of tomorrow: big data for energy transition, for smart industry, for life, and for security. At an aggregated level the program contributes to the 21st century overarching objectives of developing smart cities and empowering its smart citizens. The program will be well connected to on-going big data initiatives in, for instance, astronomy and digital humanities. The ambition of the COMMIT2DATA program is fourfold.

- A public-private research and innovation program with national scope and focused on big data, bundling data science expertise, resources and funding in the private, public and government sectors, that furthers the joint ICT expertise development in data sciences and has explicit emphasis on valorization and dissemination.
- Durable contribution to the human capital agenda in the field of ICT sciences, and in particular data science. Training of versatile data scientists – *specializing in computer science, statistics, creative processes, and technology; and with strong awareness of business and societal context* – is urgently needed for the Netherlands to be an attractive data science country to talent, companies and investors.
- Securing of earlier (public) ICT and big data investments, including the furthering of valorization results of the COMMIT research program in the above mentioned four sectors and with a strong contribution from creative sector approaches such as the use of social media and co-creation. Enhance the dissemination strategy as developed in the "ICT doorbraakproject big data" by absorbing recent ICT and data science results as so to stay in the forefront of the competitive edge. Especially in data science the cutting edge renovates every two years; constant upgrading will be necessary for at least the next decade.
- Advancing of the COMMIT2DATA results and ways of working in sector-specific regional anchor points such as data science,

valorization and dissemination centers. This also creates circular innovation ecosystems, where data science talent joins innovative industries and vice versa. At the same time leveraging Dutch excellence in data science for obtaining a leading position and thought-leadership in the European PPP on Big Data.

The results that COMMIT2DATA delivers – in addition to top-class science and reinforcement of the Dutch top-5 knowledge position in data science – can be concisely summarized as follows.

- 75 Valorization results which establish tangible transfer of cutting-edge data science research and knowledge in the form of pre-competitive "golden demonstrators".
- Dissemination of data science and state-of-the-art knowledge by interaction with 400 large and SME companies, of which at least a quarter engaged in hands-on workshops using company-specific ideas and data.
- 150 Researchers trained at Ph.D. level in use-inspired public-private collaborative data science research and innovation projects. Over 1000 students involved in projects at Master level in Academica and Applied Universities (HBO) in disciplines relevant for data science.
- 300 In-company R&D personnel who collaborated with counterparts in academia on cutting-edge data science, stewardship and technology.

Program Lines of COMMIT2DATA

In order to realize the ambitions and results mentioned above, COMMIT2DATA has been structured as three integrated program lines: a pre-competitive use-inspired research program line, a valorization sprint program line, and a dissemination program line.

Program line 1: Use-inspired research
COMMIT2DATA's use-inspired research is the program's main line. It aims at high-tech and high-science academic research impact on companies delivering advanced technology and services. This program line brings together the data science, stewardship and technology challenges from four major economic sectors and societal challenges. These sectors are energy transition, smart industry, broader life sciences & health applications, and security. The challenges have been formulated by teams of company and academic stakeholders in each sector. They are included compactly in the KAI ICT (ICT Roadmap) 2016 as sectorial breakthroughs. The context, background, anticipated impact, and required action of these research challenges have been developed in close collaboration with representative organizations from each of these sectors.

Chapter 2 of this white paper provides the details per sector. Chapter 3 describes the scientific challenges that are common to the sectors, and finally Chapter 4 explains how research projects of the COMMIT2DATA program will eventually be formed by combining data science research challenges across different sectors. These joint academic-company research projects will typically run for 4 to 5 years and provide the main basis for developing new data science knowledge. Direct valorization will be generated with founding private project partners in the four sectors.

Program line 2: Valorization sprint
The COMMIT2DATA valorization sprint program line provides the vehicle for joint research and transfer of data science results to companies and organizations

beyond those that are founding program partners and beyond the four founding sectors. Such partners see opportunities for valorization of science results achieved in one or more projects of COMMIT2DATA partners. They jointly develop an activity that leads to a "golden demonstrator" that factors in fidelity, product, usage and market considerations. In this way, COMMIT2DATA projects serve as a catalyst for focused demonstration and pre-development projects in a wide range of sectors and businesses. This model has shown to be successful in EIT ICT Labs (now EIT Digital), the STW valorization program "Take off", and COMMIT's over 50 valorization projects.

Valorization sprints aim to attract companies to COMMIT2DATA research as early as possible, and thus provides dynamics among the participating companies. The sprints result from open calls, and are typically run for a relatively short period of, say 6 to 24 months. While activities in valorization sprints themselves are still in precompetitive development phase, they aim to eventually step into the business development phase, where a commercializing company or start-up fully takes over. The effect of valorization sprints is a rapid transfer of high-tech and high-science results into technology and services-focused companies and organizations across a maximally broad spectrum.

Program line 3: Dissemination

Finally, the COMMIT2DATA dissemination program line delivers big data knowledge to an economic and societal audience that is as broad as possible. This program line aims explicitly at organizations, SME companies and other stakeholders that are owners and users of big data, but do not have the objective and/or knowledge to develop high-tech or high-science services themselves. Quite different from the traditional dissemination model of broadcasting result, the COMMIT2DATA dissemination program line moderates hands-on interaction between COMMIT2DATA (knowledge and business) partners and external stakeholders using concrete big data sets.

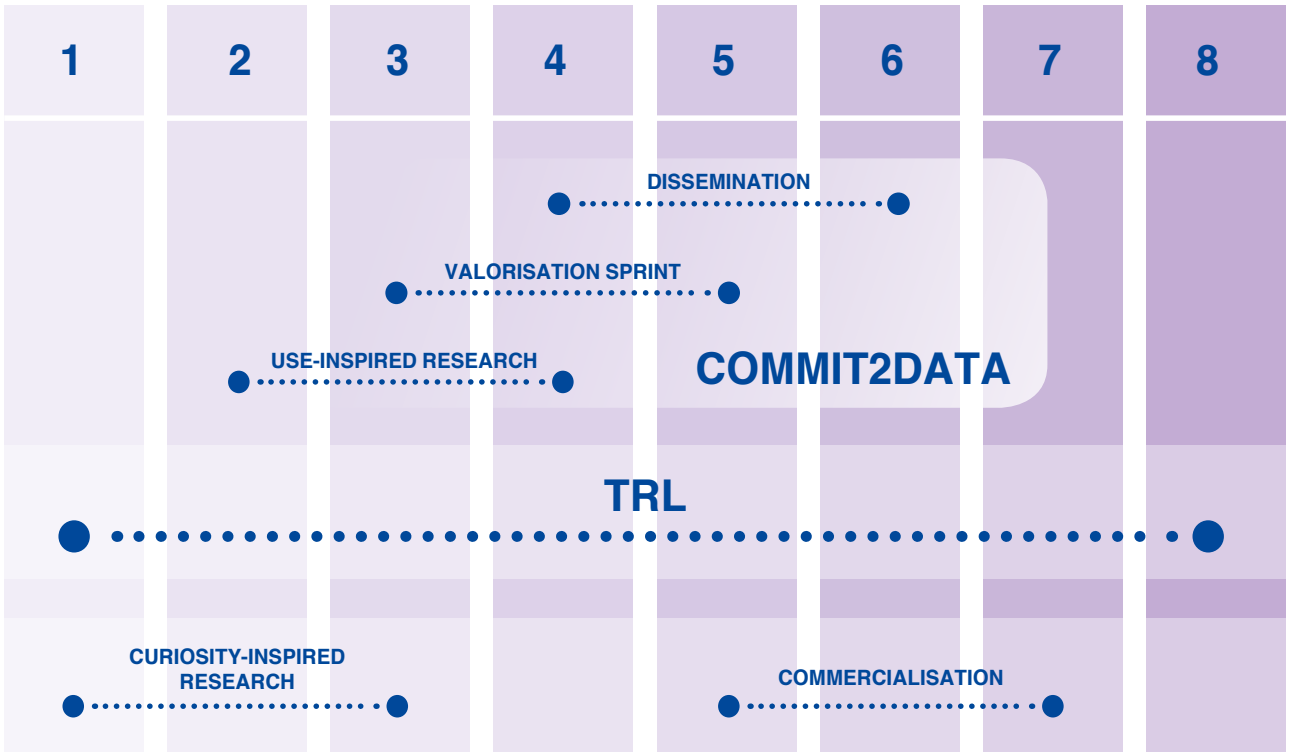
Awareness meetings, public demonstration events, and in particular data factories – *a concept developed in the "ICT doorbraakproject big data"*— will be organized to efficiently disseminate cutting-edge knowledge on big data. Chapter 4 has more details on the valorization sprint and dissemination program lines.

TRL Position of COMMIT2DATA

The three integrated program lines of COMMIT2DATA can be positioned in relation to each other using the technology-readiness level (TRL) diagram below. The three program lines are sequential in TRL level with a significant overlap. In terms of phasing in time, the program lines take place simultaneously because valorization sprints and dissemination also build on and connect to earlier initiatives.

In the diagram the position of curiosity-inspired research projects is also shown; such research is typically carried out in academia without or with little involvement of companies. It forms the foundation of on which PPPs such as COMMIT2DATA can successfully be developed, and contributes strongly to the fundamental expertise development in the field of data science in its own right. On the other side of the spectrum at the higher TRL levels, commercialization is shown, which is typically carried out in companies without or with little involvement of academia and government.

TRL POSITION
OF COMMIT2DATA



Chapter 2

Big Data in Four Innovative Economic Sectors

Big Data in Four Innovative Economic Sectors

COMMIT2DATA brings together challenges facing big data research across four major economic sectors and societal challenges. In this chapter we discuss big data in broader life sciences and health-related applications, energy transition, smart industry, and security, respectively. For each sector, contributions are delivered by representative organizations. DTL (Dutch Tech Center for Life Sciences) coordinated the contributions for big data across the life sciences & health; the TKI Urban Energy coordinated big data in the energy transition; the smart industry coordination team delivered the input for big data in smart industry; and HSD (the Hague Security Delta) coordinated the contribution for big data in security. Contributions to this chapter have been coordinated with the respective Knowledge and Innovation Agenda's in each sector.

Each contribution first describes the economics, societal and technological developments as well as the potential for big data in the sector. It then continues to give a concise characterization of the required data science breakthroughs specifically needed in the sector. Finally, the distinctive properties of the big data in the sector are summarized; these properties and associated data science challenges are the foundation for the scientific research agenda – *see Chapter 3* –, and the COMMIT2DATA program structure – *see Chapter 4.*

2.1 BIG DATA FOR LIFE[2]

Sectorial Needs and Challenges

The design of solutions to 21st century societal challenges in human health, healthcare and nutrition, and to the sustainable production of our food, feed and energy, requires innovations and novel businesses based on life science R&D in its broadest sense. A diversity of on-going programs show how tangible the public-private collaboration is in these sectors. Examples are Onco-XL, CTMM-TraIT (TransMart, Open Clinica), NFU Data4Lifesciences initiative, Parelsnoer, Centre for Personalized Cancer Treatment (CPCT) and Philips' Health-Suite Digital Platform in LSH; Breed4Food and TIFN in Agri&Food; Virtual Lab for Plant Breeding, Seed Valley and "Tuinbouw Digitaal" in Horticulture; and BeBasic in Biobased Economy.

Life-science research requires new data science, stewardship and technology solutions to enable ground-breaking discoveries, precision interventions and economic acceleration in personalized health, nutrition, crop & livestock breeding, and biotechnology. The aim is the fundamental understanding of biological complexity based on measuring and modeling "biological and medical information" processes at the spectrum from molecule to organism(s):

- molecular (e.g. genetic, proteomic, metabolic pathways) and cellular scale. In this case measurements are often made with advanced technologies in well-controlled research and/or clinical labs specialized in human, animal, plant or microbial biology;

[2] This section is based on the Knowledge and Innovation Agenda ICT 2016; Life sciences across the agenda's: Life Science & Health; Agri&Food 2016; Horticulture & Starting Materials 2016; and Biobased Economy Agenda 2016; and creative research areas from the Knowledge and Innovation Agenda Creative Industries 2016

- system level, where data are gathered across larger dimensional scales e.g. at tissue, organ or (sub)organism level, enabled by advanced imaging technologies. (e.g. 3D-dynamics of cancer organoids imaged with novel light-sheet microscopy);
- organism level of physiology, lifestyle and interaction with the environment, in processes such as health care, crop and livestock breeding or industrial biotechnology. Here measurements are often far less controlled as they are community-contributed, self-performed, or executed under harsh circumstances. Examples are measurement of blood pressure, lifestyle including the quantified-self, soil fertility, and fermentation conditions.

Precision intervention strategies are pursued at the molecular or (sub)system level that lead to the desired perturbation of (the regulation of) biological pathways and tissues. At the same time we seek handles at the organismal level to offer personalized health (e.g. handles for personalized cancer treatment, smart-watch applications for medication management), to steer crop performance (yield, quality, pest resistance, draught tolerance), or the production of specific biobased chemicals (e.g. in synthetic biology).

High-end data generation technologies are being used routinely across the life science sectors, where studies are performed on massive numbers of samples and cohorts, collected and stored in "biobanks", seed banks, livestock sperm banks and other bio-sample collections. Such data have been carefully annotated and stratified with respect to clinical or other relevant outcomes. Increasingly, also imaging information is available that allow the integration of molecular or cellular information into larger and functional structures, enabling the link between genotypical and phenotypical information. In addition, a flood of phenotypical sensory devices is used to collect information on the processes above.

This huge diversity of data gathering and analysis is crucial as we currently face the challenge of unraveling complex poly-factorial processes, for example, the prevalence of Alzheimer in severely obese people, caused by an imbalance of multiple genetic and lifestyle factors. Evidence-based approaches such as precision breeding and precision farming require the understanding of very complex systems with thousands of interdependent variables. Unfortunately, the actual possibility to intervene in biological and environmental processes for the improvement of health, agriculture, food and environment are very limited indeed, and are still based on rather simple, mechanistic and deterministic principles. Current data-driven discovery provides many explanations about complex biological systems, but also reveal why many current interventions fail, and how interventions in the future will have to be of much higher precision and personalized.

In the agro-sector, precision breeding strategies require the combination of omics-level and phenotypic properties measured through e.g. next generation sequencing and 3D spectral imaging with environmental data assembled in the greenhouse or field. Novel (bio-)chemical design strategies (for pharma, agribusiness or foods research) are required based upon data analytics at the chemical and biological pathway level to design more specific and precise bioactive chemicals in drugs. In human translational health research computational systems need to enable the combined analysis of data derived from medical devices and sensors (patient and citizen data from wearables to domotics), with personal omics data. Here, adding information derived from imaging at whole body scale, and (sub) scale enables bridging the molecular and phenomenological world.

These combinatorial analyses will roughly take place at two timescales: (a) a short (say 48-hour) timescale for diagnostics and treatment plans

based upon fixed integrated diagnosis and omics measurement regimes. And (b) longer timescales for research and innovation. Research and innovation in integrated (medical) devices, imaging and microscopy, and high throughput omics techniques at the longer timescale will steer and be dependent on short-timescale molecular and (patient) data generation.

Relevant data combinations are used for behavioral modelling, comparison, prediction and coaching, and eventually become powerful evidence-based guidelines for use in personal health strategies. Molecular and physiological data can again be enriched with data gathered in socio-economic studies or from social media. The unprecedented power of expert and lay person blogs and social media trending are increasingly recognized by life scientists and companies delivering personalized technologies, in line with P4 medicine approaches (Predictive, Personalized, Preventive & Participatory medicine). Building "system" models across dimensional scales, from molecule, to cell, to tissue, to organ, to whole system, and linking to observable, objective and quantifiable outcomes is an essential component in obtaining meaningful and implementable applications. The effectiveness of services based on data for life – *such as personalized health services* – relies on the (user) commitment to share information, contribute knowledge and experience. To that end, disciplines from the creative industries dealing with user-centered design, adoption and ethics are pivotal.

Data Science, Stewardship and Technology Challenges

Viewing the broader life sciences and health-related applications from the properties of the massively collected data and the ensuing goals of data science gives rise to the following challenges and angles for research.

- The broader life science sectors generate huge amounts of heterogeneous data at both micro- and macro-level. Genomic data alone outpaces current storage solutions, with data that is highly distributed and generated in many independent and complex formats, including spatial and temporal information. Advanced imaging and microscopy information at system, organ, organoid and cellular level is becoming abundantly available and needs to deal with huge amounts of data (e.g., digital pathology systems and light-sheet microscopy). Such information must be accessed from distributed systems, and incorporated in research, product and service development, which puts specific requirements on the capturing and validation of data design (experimental design), data annotation and data processing.
- Management and analytics of unstructured and structured data types, ontologies, versioning, and annotation of data sets over time is of imminent importance for sustained knowledge discovery in novel datasets combined with core legacy data. This also opens up new avenues for data normalization and error correction techniques based upon advanced statistics. In addition, computational simulation and mathematical modeling approaches are required targeting the dynamics of complex biological systems at multidimensional scales.
- The systems-level approaches drives life science research forward to integrate the heterogeneous data generated at multiple scales in experimental set-ups, hospital instrumentation or field studies in order to explain the complexity of the biological systems. The heterogeneity of the data reaches as far as data generated through quantified-self devices, home equipment, patient blogs and social media. Likewise camera-equipped drones and GPS-devices for precision measurements of soil aim to determine environmental influences on crop performance in smart farming.

- Data cannot be repetitively generated, and valuable datasets need to be secured for future re-use. Also the need to share and link specific datasets across institutions, companies and public-private partnerships call for secure mechanisms of data exchange, well annotated with proper metadata, and for the development and implementation of robust international standards. Novel opportunities include the "compute visits the data" approaches that increasingly uses aggregation/conclusion algorithms and models that build conclusions on serial or parallel analyses of distributed and sometimes even encrypted data sets. The FAIR initiative (http://www.datafairport.org; data must be Findable, Accessible, Interoperable and Re-usable for machines) that strongly relies on advanced semantic technology and mapping services provides an opportunity for the Dutch companies and institutes to play a key international role. The FAIR principles already have major impact on, for instance, the interoperability strategy of ELIXIR (European bioinformatics infrastructure), US-based BD2K program and the Global Association for Genomics and Health (GA4GH). The FAIR data approach can be valuable in sectors outside the broader life sciences.
- Combining and connecting data across resources and across organizations brings strong data security challenges related to privacy issues (e.g. patient information) or intellectual property (e.g. industry). Related to this topic is the willingness of (end) users to provide phenotypical data. Reliable and consistent (user) engagement models and user-centered design approaches – *for instance aiming at wearable solutions* – are needed to support intrinsic motivation to supply quantified self-data and enrich data sets. Many of these aspects are not unique for the life-sciences sectors, and can be worked on in close alignment with other sectors.

Referring to "Two dimensions of COMMIT2DATA" in Chapter 1, we see that data science for life is driven strongly by the volume, velocity, and heterogeneity of the data. Privacy, access and longevity of collected data are also primary considerations. In terms of impact, the sector is characterized by gaining insight via data-driven hypothesis testing, as well as economic and societal value through more experienced scientist and engineers being able to interpret patterns in data and contribute to sustainable health, healthcare, nutrition, food and energy solutions.

2.2 BIG DATA FOR ENERGY TRANSITION[3]

Sectorial Needs and Challenges

Large-scale application of renewable energy resources, digitalization and electrification of our society, operating in a cost-effective and resilient infrastructure can only be accomplished by means of intelligent networks: the smart grids. Data will play a crucial role within the changing business models and market mechanisms to guarantee access to energy. The traditional model, solely driven by demand, can now be complemented with a supply-oriented approach and mechanisms for storage and conversion. This leads to a more sustainable energy system with less emissions, lower prices,

[3] This section is based on the Knowledge and Innovation Agenda ICT 2016; Knowledge and Innovation Agenda Urban Energy 2016; and creative research areas from the Knowledge and Innovation Agenda Creative Industries 2016.

bottom up commitment and increased resilience through multiple methods for flexibility and availability. The top sector Energy was erected in anticipation of these developments, and it has been an instrument in bringing together stakeholders in public-private partnerships setting the agenda for the energy transition.

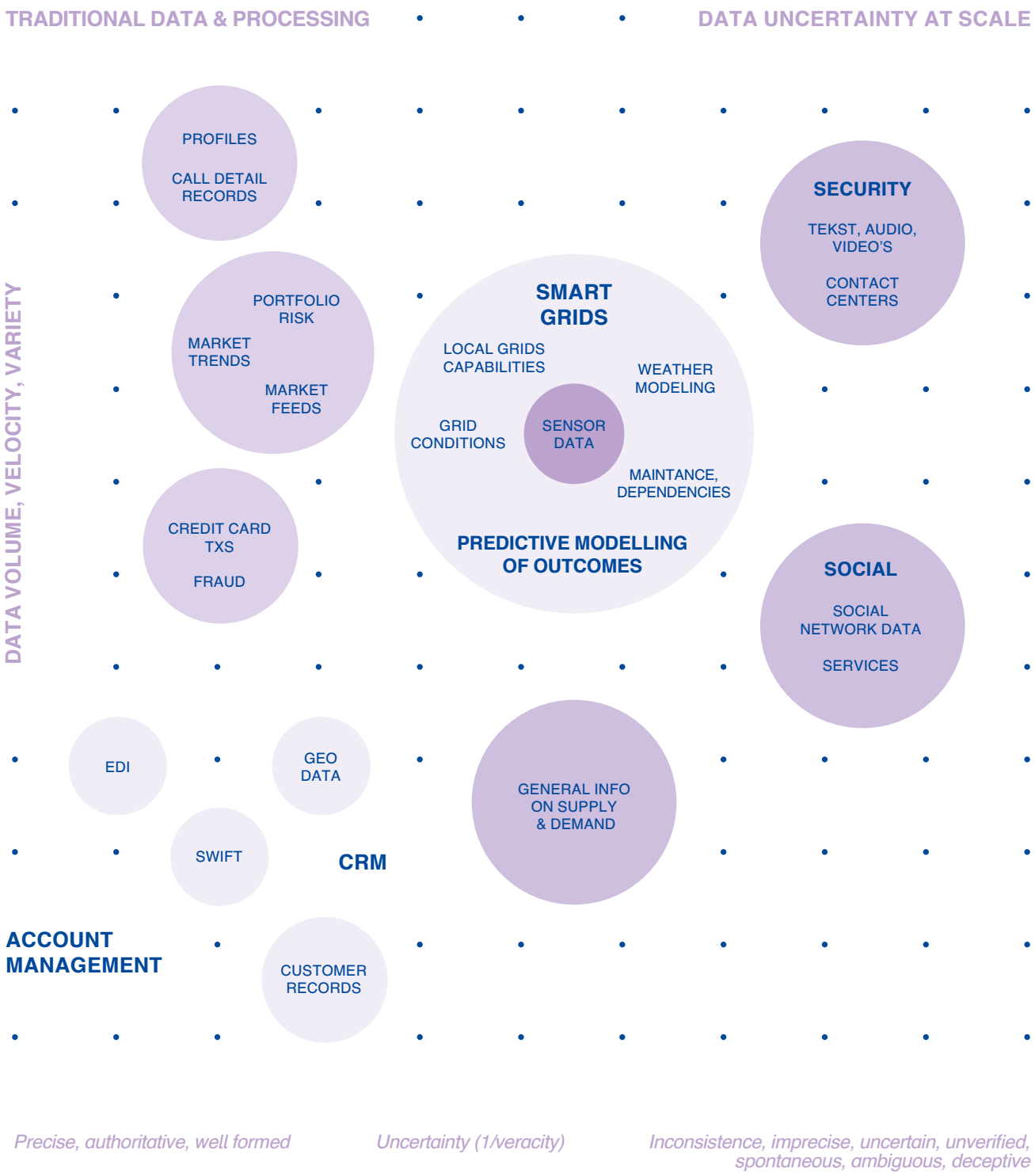
The energy transition is a process through unchartered territory in the midst of many complicating factor. These include, for instance, geo-political changes, the threats of boycotts in gas delivery, developments in the Middle East, conflict of interest even within Europe, and the fact that alternatives for local cities and countries are often not more sustainable than their current dependencies. The energy transition is therefore far more than a technical issue, it needs the creation of a "new order" in the supply of energy. Deployment of smart grids is needed to accommodate the new order and cater for entirely different roles than today's system was designed for. For instance, at this moment most stabilizing controls are mounted at the large scale generation edge of the grid. Meanwhile new, sustainable generation will grow at the consumer's edge where the grid lacks stabilizing mechanisms. Smart grids can also reduce the infrastructure capacity needed to meet local demand by more than 30%, hence cutting investment costs.

In a way, smart grids are partially the implementation of the industrial internet that has become an integral part of many systems. Like in other sectors, integration is a dominant factor, both central and de-central. The driving factors for integration, cyber-physical systems and the internet-of-things (IoT), will dominate this environment, inevitably producing vast amounts of data. This energy network data will not exist in splendid isolation. It will be combined with many other data sources, like geo-information, weather data, maintenance reports, condition incidents, and consumer-reported social data. Predictive analytics will bring this data to life and apply it actively in day-to-day operations, customer relations, and in long-term decision making. Companies will have to learn to use data-driven decisions as rules for success or even to survive.

The number of organizations that will be affected by such use of big data in the energy sector is significant. Examples are distribution system operators (DSOs) and transmission system operators (TSOs); energy traders and retailers; system integrators; telecommunication companies; energy users and suppliers from consumers (prosumers) to responsible parties in smart industry and horticulture; institutional players like banks and government that need to help to overcome regulatory, financial and organizational hurdles. Many of these organizations will have to do much more than optimize current operations; they will primarily need to investigate and implement new business models based on big data. If they do not do it, others will.

In the energy sector – *and in fact the utilities sector at large* – we will see many forms of data, with different attributes and properties. Nevertheless, some typicality's exist for this data. The network data itself is real time, uncontrolled, de-centralized, and involves many actors. In fact, the number of actors will only increase, also outside the traditional energy sector. Data has a strong streaming character as the network operates 24/7, is heterogeneous due to the different hierarchical levels at which data is collected, and data is often measured in harsh environments yielding uncertainties. Some data comes in very large volumes, centrally generated, other data is locally produced. Much smaller data sets also exist that need to be integrated such as historical records.

DIAGRAM ILLUSTRATING THE WIDELY VARYING CHARACTER
OF DATA IN THE ENERGY (UTILITIES) SECTOR



On the relatively short term, big data that is acquired, combined and analyzed in the energy sector will be used for operating (early versions of) the smart grid, for new data services, and for asset and workforce management. Already now the benefits of these techniques have been demonstrated in smart grid living labs yielding a considerable societal business case[4]. New services will be offered to customers for instantaneous and future energy management based on big data. In the long run, big data will not only be used for operational systems, but also for knowledge and cognitive systems. For an energy system design in which we can place out trust and wherein democracy rules as in real life. A very important choice that needs to be guided by many factors including big data, is the architecture of the future system. Will it be compute-centric or data-centric, whether or not it will be agent-based, whether it will be a (data) centralized or decentralized, or a combination of all these.

Finally, the transition of the energy sector is also a matter of user perception and behavior: how do users engage with the technical systems and how do we avoid users' information overload? User-centered design approaches are therefore integral part of the use of big data for energy transition.

Data Science, Stewardship and Technology Challenges

Viewing the energy sector from the data properties and the ensuing goals of data science gives rise to the following challenges.

- Big data has opened a new frontier for data stewardship efforts in the energy sector. The future is not going to be as straightforward for data stewards as dealing with conventional structured transaction data. Collections of data have become highly variable and may include a mix of structured and unstructured data types: transaction data, system and network log files, information from sensors, internet search records and text-based social networking data, real time text and video, from all kind of sources in a geographical and topological context. Such data often comes from external systems, adding another complicating factor for data stewards, as they cannot exert any control over the quality and consistency of the information as it is being created. It needs to be known wherefrom observations came and when. Inserts, changes and deletions must be accounted for, in real time, in sub-seconds. Finally, also the interoperability and underlying architectures of such heterogeneous data formats – *including open data* – across organizations is essential for the sector.
- Management of the many classes of assets in the sector. Asset management is increasingly going to be determined by the collected big data. Hence, data science should strongly be directed at using big data for decision support operations, including problem management and network load management. Decision support relies strongly on predictive analytics, which demands data analytics to rise above the level of simple pattern finding by correlation but rather address the issue of finding root causes, i.e. causality.
- A particular issue in the energy sector is scalability from the perspective of costs of transport of energy versus data. The introduction of cyber physical systems, the transition to smart grids, and the use of data analytics in energy systems should not affect the sustainability. Local intelligence, data selection, reduction and compression techniques may be needed to avoid overly large energy costs in local and remote processing of energy-related data.

[4] https://www.dnvgl.com/technology-innovation/broader-view/sustainable-future/vision-stories/power-matching-city.html.

- The energy sector is also specific in its privacy and security issues due to the fact that energy can be considered a critical infrastructure with an extreme societal relevance. Collective safeness and individual freedom need to be carefully balanced. Design and deployment of advanced analytics need to be based on privacy and security by design approaches; yet it is important to realize that the current old infrastructure enhanced with modern equipment like routers and switches created a situation where security by design is virtually impossible. A possibly viable option – *yet at the same time one of the toughest challenges* – is to perform advanced analytics over cryptographically protected data such that organizations can effectively anonymize data before sharing the information.
- User-centered analytics and design have demonstrated that energy transition in household-consumers and professional energy users can be achieved if user perception and awareness is properly addressed. Big data-based system designs are therefore desirable that achieve intrinsic motivation for change.

Referring to "Two dimensions of COMMIT2DATA" in Chapter 1, we observe that data science in the Energy Transition is driven by many factors simultaneously. The volume, velocity, and heterogeneity of the data is currently fairly manageable but is expected to exponentially grow once smart grids come into operations at the large scale. The quality of the data is a major issue due to the lack of control over system and data sources, which makes reliable decision support a big challenge. Theft, privacy, access and longevity of data is another major challenge for the energy sector. Finally, the impact of the data is in economic and societal value of the reliable, sustainable and affordable energy system.

2.3 BIG DATA FOR SMART INDUSTRY[5]

Sectorial Needs and Challenges

The concept of Smart Industry reflects the fourth industrial revolution with so-called cyber-physical systems, emphasizing that everything is now digitized and digitally interconnected. This technological innovation will lead to business and social innovation. Traditionally, a product was designed, the bill-of-materials compiled, the parts bought or manufactured and the product was subsequently assembled and sold. Today, however, we see ever-increasing complex products and associated services coming out of value constellations of suppliers, manufactures, and brand owners. In the near future the value chain is yet to experience another shift with the introduction of product personalization and new technologies, such as additive manufacturing. Products themselves are also becoming increasingly intelligent, with embedded computing inside. This not only enriches the intended functionality of the products, but smarter products will also be able to monitor their intended and unintended use, and failure in all phases of the life cycle.

Action line 12 of the 2014 Smart Industry Action Agenda[6] – *titled "Big Data - Big trust"* – emphasizes the importance of big data from two perspectives. First, the value chain will become increasingly integrated. Companies involved in manufacturing will increasingly depend on each other's half-products and the accompanying

[4] This section is based on the Knowledge and Innovation Agenda ICT 2016; the Smart Industry Knowledge Agenda 2015; Smart Industry Action Agenda 2014; and creative research areas from the Knowledge and Innovation Agenda Creative Industries 2016.

[5] http://www.smartindustry.nl/wp-content/uploads/2014/11/Smart-Industry-actieagenda-LR.pdf.

-

production and quality data. Ideally such data is highly structured, but common practice is that also a lot of information is scattered and incomplete, informally formulated, and available in a diversity of carriers from text to visual information. Furthermore, on both sides the value chain will be extended towards end users and customers. Personalization and “one-of” production stimulates the transition to data-driven servicification of manufacturing industry. User interaction for specification of complex products is a major challenge. After production and deployment, the data collected during use will be fed back into the value chain. This data will be measured by the product itself, but also end-user feedback in the form of messages, photos, and comments on social media. Even human emotions while using the product or opinions mined from social media on a particular product will become part of the value chain.

Second, automation of the production processes themselves will be taken to the next level. New generations of manufacturing machinery will be highly robotized and integrate a diversity of innovative cyber-physical systems. These machines produce continuous streams of data that will be stored, analyzed, and put to use so that production becomes intelligent, flawless (zero-defects) and self-learning in a wide range of environments and conditions. As the size and complexity of man-made manufacturing systems is already large and it is increasing by the day, the results of data analysis will also be used for business intelligence and decision support in designing and operating fallback mechanisms, maximizing production availability, and maintenance or software updates of complex system-of-systems. Production processes themselves will also be more and more often monitored. Continually real-time information will be provided about environmental parameters such as air quality. The massive streams of real-time data from multitudes of sensors, devices, and instrumentation and other sources makes is necessary to conduct information mining and visualization to assist operators in the manufacturing industry.

Hence, the days that ICT in industry encompassed merely the storing product and manufacturing information are in the past; in tomorrow’s industry life cycle, quality, usage and other relevant and highly diverse data of every individual product and component in the production chain will be stored, used, and shared across the value chain. This will enable production at higher yield, higher quality, lower costs and increased flexibility.

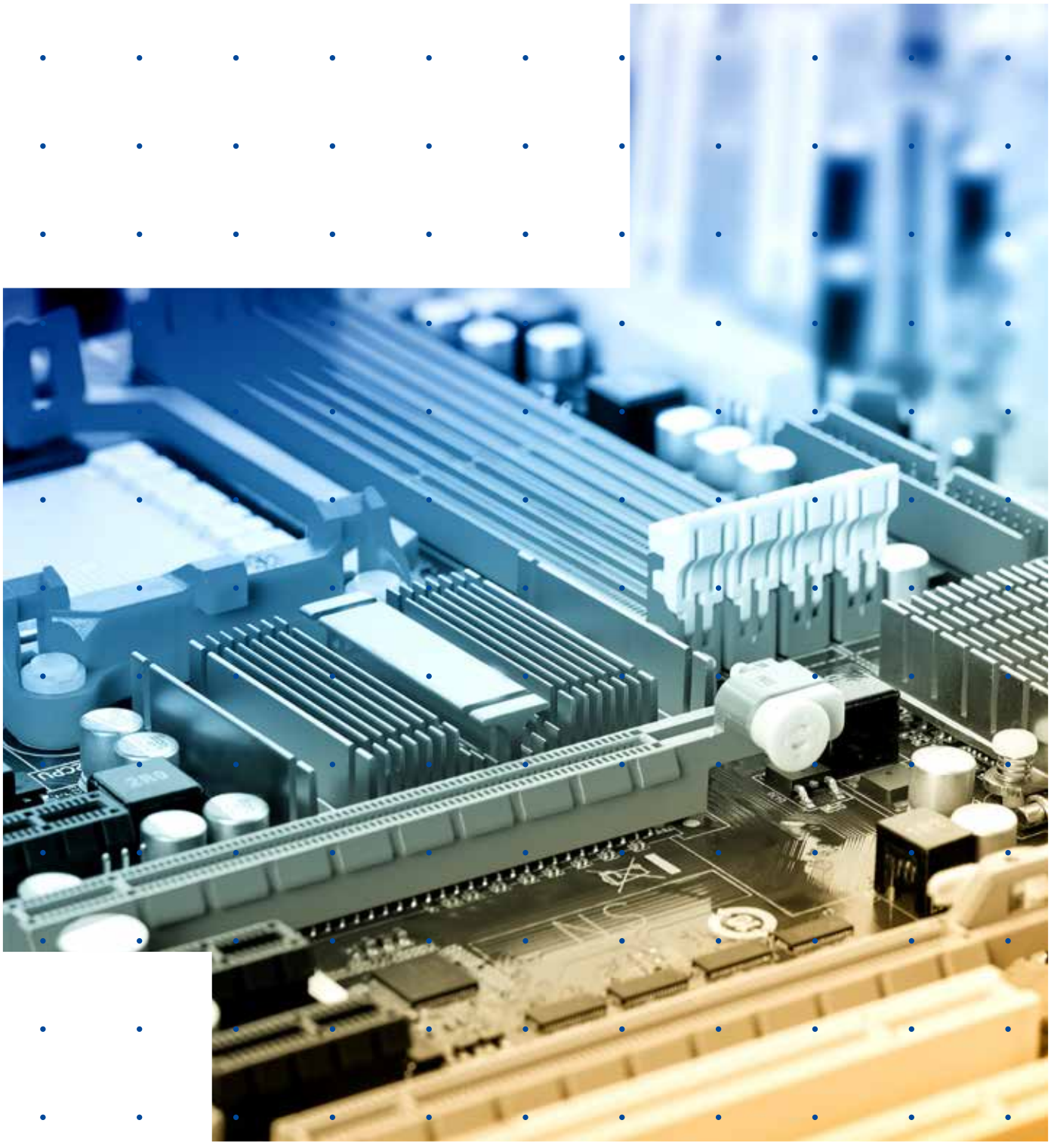
In 2015, the implementation of the Smart Industry agenda began with establishing a number of so-called Field Labs. Action line 6 of the Smart Industry agenda emphasizes the need for strengthening R&D in the Field Labs. The Field Lab “Region of Smart Factories (RoSF)” focuses on zero-defect production and “first-time right” product and process development. This Field Lab, as an example, requires extensive sensory and video monitoring and analysis of big data in real-time. In “Smart Dairy Farming (SDF)”, real-time analysis of sensory data is required for monitoring milk production, as well as sharing of data across the value chain. The securing of the exchange of data across the complete value chain is a particular focus of the Field Lab “Secure Connected System Garden (SCSG)”. A final example is “Ultra personalized products and services (UPPS)” that focuses on innovative applications of collected data, such as clothing-integrated sensors in the design and care sectors. Product-service design principles will need to be applied to achieve end-user engagement and adoption. Thus opening the way for on-demand production of high tech products and services.

Data Science, Stewardship and Technology Challenges

Viewing smart industry from the properties of the massively collected data and the ensuing goals of data science gives rise to the following challenges.

- The software, hardware, business processes and end user will generate huge streams of heterogeneous data, from simple sensor data to complex 3D video streams of production processes or products themselves. In some cases this data will need to be utilized directly by operators to improve quality or to control processes otherwise. Furthermore, data analysis should factor in that some data might be missing or is unreliable due to sensor failures and system transients. Dealing with these data properties requires a high level of cognition of the overall data analysis process, and sufficient interaction methods with the human operator.
- Smart Industry is characterized by value chains, involving multiple actors who are sometime competitors, and heterogeneous controlled and uncontrolled data. Finding the right information – *that can be accessed and trusted* – is not trivial, and requires novel search methods on the "industrial internet".
- Standardization has a significant impact on the practical introduction of big data methods in Smart Industry. Standardization, interoperability of proprietary and open data, and first-mover advantage requires the contribution of good research and ideas to European and global standardization initiatives by Dutch parties. Data standards should cover typical industrial parameters, such as quality, delivery guarantees, ownership, authenticity and integrity of the data.
- Data will need to be shared in order to make it possible to share business. The creation of new services and businesses in efficient chains based on big data therefore requires a high level of trust between the parties in the value chain. Trust-by-design could turn into distributed (peer-to-peer) trust solutions, or into industrial safe houses. In any case, such solutions must be based on secure storage, tracking, tracing, and sharing of data beyond the current concept of trusted third parties, as the constellation of involved parties in smart industries consists of complex and variable interactions of services, components, and data. Also the constellation of partners is not static but subject to changes. Newcomers must be accommodated in becoming part of the trusted community, the trust of companies who have left or were out casted should be revoked.
- Finally, even today it is already difficult to be off-the-grid. With the large scale introduction of big data in smart industry, all products will become on-line as part of internet-of-things. Data from sensors in products and social networking sites will enable highly accurate tracking and tracing of products and their use. Smart industry therefore demands solutions to avoid infringement of privacy while not limiting the power of the use of collected data. On demand production asks for user-centered design methods fully incorporated in the smart industry chains.

Referring to "Two dimensions of COMMIT2DATA" in Chapter 1, we observe that data science, stewardship and technology in Smart Industry is driven by hard aspects of data, and by trust & privacy issues. The volume, velocity, and heterogeneity of the data will greatly expand as these are driving factors in the smartness of industry. The quality of the data is an issue due to the difficult (industrial) monitoring conditions and the increasing use of consumer product data. Trust and privacy associated with big data require a lot of attention as the success of the multi-actor value chain is going to be critically dependent on them. Finally, the impact of the data in smart industry is mostly in economic (business) value but the impact on social innovation is also going to be significant.



2.4 BIG DATA FOR SECURITY[7]

Sectorial Needs and Challenges

"Security creates the conditions for societal stability and economic development. Without that stability, Amsterdam would not have been able to grow into a world player in the internet exchange business, and the main ports Rotterdam and Schiphol would not be able to fulfill their hub function"[8]. Effectiveness in guaranteeing security can only be achieved by the availability and quality of information based on big data from a multitude of sources and sensors, be they physical or virtual, machines and/or human. Security technologies and services based on big data provide ample economic opportunities for the Netherlands. Furthermore, the government is a dominant player as regulator, enforcer but also as end user for security products, services and innovations.

Security concerns government, companies, citizens and society at large. Due to an increasingly globalized world, the concept of security has become much broader and more interwoven with other areas in society like living, industries, transport. Smart cities must be secure cities almost by definition. Security therefore has become a multi-party challenge largely driven by data. Connections between parties are sometimes short and temporary in nature. Indeed, with the increasing number of parties having a role in security, this may well become the standard. This puts additional demands on the (data) processes, structures and systems designed to connect sensors and actors in networks and chains quickly, on an ad hoc basis, and yet still in a reliable manner. Furthermore, "where security issues in general become more complex and dynamic, this applies even more to the digital domain. Cyber threats are developing super-fast and it is common knowledge that governments, companies and people are insufficiently equipped to deal with these threats."[9]

Increasingly individuals, companies and societal organizations are interested in playing an active role, e.g., providing security around the railways, responding to aggression on public transit, coming together to provide security in communities and urban districts, preventing nightlife violence or ensuring collective security on business parks, and industrial estates. The "man in the street" enters the security domain as a sensor, often via textual, audio and visual content spread via social media. But also a broad range of autonomous (audio-visual) monitoring, surveillance and detection sensors are emerging such as unmanned satellites and UAVs. Between these stakeholders, resources and data need to be shared, such that situational awareness can be developed and coordinated, and effective action can follow.

Security systems are increasingly functioning in chains and networks, using data from many sources to create real time intelligence, resulting in better responses to incidents and crises and more effective criminal investigation and public order management.

7 This section is based on the Knowledge and Innovation Agenda ICT 2016; Roadmap HTSM Security 2016; Ministry of Security & Justice: "Big data, veiligheid en privacy", 2014; and the National Cyber Security Research Agenda II, 2013.

8 Rob de Wijk: National Innovation Agenda for Security 2015.

9 Roadmap HTSM Security, 2016.

10 Ministry of Security & Justice: "Big data, veiligheid en privacy". 2014.

-

Examples in safety range from sensors monitoring the stability of dikes, Google that is able to monitor flu-epidemics using its keyword search data, the analysis of the number of burglaries in different geographical region. Networked security requires data, technologies and networks to co-evolve towards systems of systems, requiring new ways of organizing and interaction beyond professional boundaries, and redefining traditional roles and responsibilities. In the digital domain, big data based solutions will help to detect, stop or avoid large-scale DDoS attacks, epidemic virus distribution, and stealthy and dormant attacks on high-value targets.

The ongoing increase in gathering information necessitates novel concepts of processing and understanding these data. However, privacy of the citizens and the workload related to interpretation of the data collected put serious constraints. Legislation and regulations have to be able to keep up, if they are not to become a limiting factor on the necessary innovation. Legal, privacy issues, ethical and administrative issues present important considerations, and in some cases limitations, on security solutions. The Ministry of Security & Justice has requested the Netherlands Scientific Council for Government Policy (WRR) to investigate the opportunities and threats of the use of big data.[10] They are currently investigating issues related to the transparency of data mining, how the use of big data will influence the effectiveness of police, data protection, and the reliability of open data. The application of big data in the security domain for shared situational awareness, real time intelligence, and business intelligence will continue to grow. Continuous research and innovations in big data science, stewardship and technology will be needed to unlock its full potential.

Data Science, Stewardship and Technology Challenges

Viewing the security domain from the properties of the massively collected data and the ensuing goals of data science gives rise to the following challenges.

- Big data is used to construct a common operational picture (COP), and directing actions based on that COP. To effectively use the exponentially growing amounts of data from different sensors and sources requires automated methods to structure, verify, sort, combine and interpret it. Algorithms for pattern recognition must be further developed. Data and sources will need to be qualified, validated, verified and reliability established to get a reliable COP. Careful considerations must be made which sources are included that will actually add value; rather than mechanically including all there is.
- Data is not only used in a reactive way to detect trends and building up a COP but also in a pro-active and predictive way. Big data and data mining techniques are needed that identify irregular, undesired and/or prohibited behavior and subsequently predict future criminal or hostile conduct.
- Maybe more so than in other domains, identified patterns must be related to causality. A high degree of transparency is needed about data-driven reasoning because results are often needed for subsequent processes like prosecution, forensics and reconstruction. Especially from this perspective, it is important to balance the part of the interpretation that is e done algorithmically by computers and the part done by human analysts.
- Also somewhat unique to the security domain is the required resilience and robustness of data analysis methods against disinformation. By definition, security deals with actors with malicious intent. These actors can

purposefully create data intended to provide false information that leads to a wrong understanding of what is going on (COP), purposefully misdirect (expected) interventions or provide "denial of service" thus obstructing and hampering security operations. Methods and technologies must be able to identify disinformation and be robust against fabricated data and data abuse.

- Collecting data needs to be "legal by design" as security organizations and companies operate within a defined mandate and collection of data can be restricted. In a controlled information world, data must be validated, stored effectively and destroyed when legally required. In the distributed information world with few controls on duplication of data, the question of ownership and destruction of information remains a largely unanswered one.
- The significant increased possibilities of gathering and storing data have led to a growing threat of data falling in the hands of criminals, foreign governments, and terrorists. Protection of data is not only a concern for the security domain but also poses a serious challenge for the other domains such as energy, life sciences, and smart industry. Access management and compliance monitoring techniques to preserve confidentiality, availability, and integrity of data according to well-defined security policies and ethical considerations on data ownership and data usage are getting increasingly complex and important.[11] "Data security is a major concern among European companies across sectors, as consistently highlighted (…) in recent years. Sound security, governance, and risk management processes need to be put in place, as big data adds legal, ethical, and regulatory considerations to data analysis efforts, and introduces new risks when data is made public or personal data is used, expanding the potential for public missteps which could bring about fines and permanent damages to the company's image and respectability."[12]

Referring to "Two dimensions of COMMIT2DATA" in Chapter 1, we see that data science challenges in the security sector are dominated by the heterogeneity of the data and the reliability of conclusions being drawn on patterns and causality, also in view of fabrication of counterfeit data by adversaries. Privacy and the legal usage of collected data are important considerations. In terms of impact of the big data, the sector is characterized by the desire to gain insight and steer action for the purpose of societal and economic stability. At the same time the pervasive role of security and the availability of big data across many societal and economic sectors leads to strong economic opportunities.

[11] National Cyber Security Research Agenda II, 2013.

[12] Business opportunities: Big Data Report for European Union, 2013, https://ec.europa.eu/futurium/en/content/business-opportunities-big-data.

Chapter 3

Scientific Challenges and Excellence

QUOTES SHOWING THE SHARED DATA PROPERTIES AND RELATED RESEARCH CHALLENGES PER SECTOR

DATA PROPERTY / RESEARCH CHALLENGE	FOR LIFE	ENERGY TRANSITION	SMART INDUSTRY	SECURITY
VOLUME, VELOCITY, AND TEROGENEITY	Finding meaning and causality Technologies for computational complexity	*" integrate huge amounts of hetero-geneous data"*	*" address the issue of finding root causes"*	*" huge streams of heteroge-neous data"*	*" effectively use exponentially growing amounts of data from different sensors and sources"*
QUALITY AND VARIABILITY	Self-learning and predictive analytics	*" data normalization and error correction techniques"*	*" data have become highly variable"*	*" difficult (industrial) monitoring conditions"*	*" resilience and robustness of methods against disinformation"*
MANAGEMENT AND PROTECTION	Interoperability and standard-ization	*" datasets need to be secured for future re-use"*	*" interoperability and underlying architectures of formats across organizations"*	*" standardization, interoperability, and first-mover advantage"*	*" validated, stored effectively and destroyed when legally required"*
	Data privacy and security	*" FAIR principle"*	*" privacy and security by design approaches"*	*" value chains, involving multiple actors"*	*" data falling in the hands of criminals, foreign governments, and terrorists"*
IMPACT AND VALUE	Storytelling and Design	*" novel busines-ses based on life science R&D in its broadest sense"*	*" economic and societal value of the sustainable and affordable energy system"*	*" economic (business) value and impact on social innovation"*	*" security creates conditions for societal stability and economic development"*

Scientific Challenges and Excellence

Most data science, stewardship and technology challenges are common to the collective of economic and societal sectors. These scientific challenges form the core of the COMMIT2DATA program, interconnecting the economic and societal sectors from a research, valorization and dissemination perspective. In the table below we summarize the data properties and research challenges (see later in this chapter) with quotes from the sector descriptions in Chapter 2.

Cross-Sectorial Data Science, Stewardship and Technology Challenges

Data science is the new and rapidly emerging scientific discipline that aims at generalizable solutions for extracting insights from data, communicating these insights to users, and safeguarding proper storage and usage of the data. One of the objectives of COMMIT2DATA is to train and deliver specialists in this field, often called “data scientists”. They seek to use all relevant, often complex and heterogeneous data to effectively convey a data-driven conclusion that can be easily understood by domain experts. Data scientists do so by integrating techniques and theories from many fields, including statistics, data analytics, pattern recognition, machine learning, online algorithms, visualization, security, uncertainty modeling, big software, and performance computing. They find interesting, surprising and reproducible patterns in data that lead to new insights that can be used to make reliable predictions.

Research – *and the ensuing valorization and dissemination* – within COMMIT2DATA will deliver solutions to the following set of data science, stewardship and technology questions underlying the table above. These data science challenges are included in the Knowledge and Innovation Agenda ICT 2016 as integral part of the Action Lines “Data3: Big Data” and “ICT One Can Rely On”. Furthermore, these questions are well-aligned with those included in the NWO grand challenge “Big Data”, and contributions to the National Research Agenda (NWA) by ICT and data scientists. A focused effort of ICT sciences is needed to push the boundaries of the today’s solutions and to address the fundamental understanding of semantics of data, computational complexity, data protection, and human information overload.

- **Finding Meaning and Causality**
  How can machines find patterns and causal relations in heterogeneous data sets, and in which way can machines learn from humans to understand these data? Isolated data are meaningless, they must be embedded in context in order to find meaningful patterns and to interpret these patterns semantically. Therefore machines must be able to deal with heterogeneous and often uncertain data including measured numbers, words, documents, sounds, images and video. Yet the number of patterns that exist in these data may be endless; algorithms will be able to intelligently select those that are interesting, meaningful, and actionable in a particular application context, mimicking human data analysis. Progress on this challenge requires COMMIT2DATA expertise in machine learning, pattern recognition for images, language and other media, artificial intelligence, content and web technology, interaction, visualization, statistics and process/data mining.

- **Self-Learning and Predictive Analytics**
  How can we create self-learning algorithms, that learn from past experiences and learn continuously as data becomes available, and that are able to make predictions on events that have not occurred before? The use of predictive analytics gives rise to new observations. How well did the predictions match the later observations? Did processes and users behave as predicted, and if not, how will the algorithm learn from its mistakes, possibly instructed by explicit or implicit user feedback. Processes underlying observations often change over time. For instance populations, preferences, consumption patterns, and performance are parameters in data analytics that are dynamic. How to deal with these continuous changes without freezing time and without

operating on outdated or legacy data? Big data collections may be rich, but rare events may not be present in the data. Nevertheless, these events may be of high relevance in case they require, for instance, specific actions of human operators. Can data science make predictions on such events (in real-time) based on inferred causal models? In order to address these challenges successfully, COMMIT2DATA needs expertise in the fields of statistics, deep learning, process mining, modeling, and artificial intelligence.

- **Technologies for Computational Complexity**
  How do we deal with the complexity of big data? The traditional approach to analyzing data was based on querying databases, relying on complete and correct answers. Finding useful information in unstructured, incomplete and partially incorrect big data requires a far more interactive mode of operation, where the user has the means to stepwise explore the data deeper and deeper into the database. Big data not only drives the development of new databases for business intelligence, process control, and data exploration. Also software innovations are needed to deal with the data explosion and the complexity of the underlying distributed computing infrastructure. Modern open source tools as Hadoop MapReduce partially address these challenges. However, these tools need to be extended to deal with the above more complex big data analysis functionalities such as learning and reasoning. Progress in new data science and technology is needed to deal with the complexity of big data; this requires contributions to COMMIT2DATA from the fields of databases, visualization, software engineering, programming languages, and computing architectures and hardware.

- **Data Privacy and Security**
  Data analytics relies on the availability of rich data containing patterns. By implication this collected data also reveals something about behavior in the digital and physical world, about identities, and about critical and sensitive business processes. How can an individual's privacy and company's trade secrets be protected while at the same time allowing for data analytics to do its work? New data science and stewardship solutions will need to be developed based on improved data anonymization and data encryption techniques. Also discrimination- or manipulation-aware mining techniques will be needed that aim to make results more fair. Such techniques will have to face the ever increasing computational power available to adversaries, but also find solutions to the undesirable overhead that cryptographic techniques impose on data analytics algorithms. Expertise in COMMIT2DATA is needed in the fields of (cyber-) security, identity management, information protection, security- and privacy-by-design, applied cryptography, and efficient algorithms.

- **Interoperability and Standardization**
  In order to find, access, exchange, and maintain big data – *including open data* – in the long run, interoperability needs to be guaranteed. Semantic interoperability emphasizes the need for interoperability of the meaning of the data. Yet, if data is collected for one purpose, how can we make sure that it can be repurposed for a completely different application? Common standardization approaches to semantic interoperability are adding meta-data, or linking data elements to controlled shared vocabularies. Linked (open) data describes a method of publishing structured data so that it can be interlinked and become more valuable. But such standards will still need to allow human understanding and interaction when exchanging data.

Optimal cooperation is needed between the cognitive strength of humans, who can understand the meaning of data, and the data processing strength of machines that have difficulties in understanding the meaning of data in the context of the everyday human world. For COMMIT2DATA to deliver progress on this challenge, expertise is needed in the field of information systems, information and system architectures, software engineering, and standardization processes.

- **Storytelling and Design**
  The ultimate goal of data analytics and prediction is to identify patterns in the data that are actionable, and leads to intellectual or economic impact and value. Efficient mechanisms and designs are needed to communicate, explain and interact with these patterns. Visualization of the results of data analysis are becoming integral part of data science challenges. But also framing results as well-designed stories or game play are essential methods for data scientists to interact with non-specialists. To that end, COMMIT2DATA needs expertise on graphics and visualization, language technology, serious playful interaction, gamification and user-centered design, and simulation and animation.

The above mentioned challenges emphasize the need for scientific progress on data science, methodologies and engineering solutions. However, as has been pointed out before, big data does not live in isolation. Its application context demands more than methodological and engineering solutions. For instance, in smart industry social sciences are important to study the impact of – *and derive guidelines from* – the effects of digitization and smartness of industrial processes such as robotic systems.

In security, legislation restricts the endless use of data for all purposes, while in life sciences ethical discussions play a similar role. Finally, monitoring and coaching applications in energy transition and health require a solid creative and design approach so that solutions are attractive and acceptable to end users. These are just a number of examples explaining the need for collaboration of data scientists with non-technical disciplines within COMMIT2DATA. In this way COMMIT2DATA delivers so-called T-shaped scientists and researchers who can communicate beyond their own data science discipline with domain experts, data owners and end users. They have knowledge and experience in methodological and engineering approaches, business and society, user-centered design and they have entrepreneurial skills.

Data Science, Stewardship and Technology Expertise

To successfully achieve progress on the above mentioned challenges, the COMMIT2DATA program builds on the core Information and Communication Technology (ICT) research groups at Dutch universities. International research assessments show that the Netherlands has an excellent academic knowledge basis in ICT. In the words of the international committee on computer science university research assessment[13]: *"Computer science in the Netherlands is a vibrant enterprise. In each department the committee saw strong evidence of excellence (…). As a country, the Netherlands remains among the top nations in computer science research, and in the absolute top in a number of sub-areas".* A recent EU study on ICT poles of excellence[14] shows that the regions Amsterdam, Eindhoven and Delft belong to the top-20 of important ICT-innovation areas within

[13] http://www.win.tue.nl/cwb/Computer-Science-March2010.pdf.

[14] http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=7140.

the European Union, partly thanks to intensive public-private collaborations.

In the field of data science, universities have put forward consolidated research and valorization efforts with substantial first money stream commitments, and will continue to expand these activities in next decade. Without claiming completeness, we mention a number of these focused initiatives. In Amsterdam ADS (Amsterdam Data Science) has been established as a regional collaboration between UvA, HvA, VU, and CWI. Together with IBM, VU and UvA develop CHAT focusing on big data and digital humanities. At TU Eindhoven, DSC/e (Data Science Center Eindhoven) has been established with close ties to the Brainport region, University of Tilburg and city of Den Bosch. At TU Delft, DDS (Delft Data Science) concentrates on the engineering of big data solutions in close cooperation with AMS (Advanced Metropolitan Solutions), Medical Delta, and The Hague Security Delta; cooperation with IBM is shaped as Collaborative Innovation Center (CIC) on Big Data. Leiden nelUniversity founded LCDS (Leiden Center for Data Science) with focus on life sciences. At University of Groningen, the DSCC (Data Science and Complexity Center) was recently established. Together with IBM, ASTRON and University of Groningen developed the ERCET initiative aiming at exascale computing in astronomy, life sciences and energy. Most universities with ICT expertise have initiated specific master programs on data science and technology in recent years, sometimes in collaboration with non-ICT faculties to underline the required broad scope (T-shape) of data scientists.

Further underpinning the excellent position of Dutch academia in data science is that TU Eindhoven leads a consortium of several Dutch universities and organizations which are organizing the European Data Forum (EDF 2016) in the Netherlands. The European Data Forum is the annual European meeting place for industry, research, public authorities and other initiatives to discuss the challenges and opportunities of big data in Europe.

Within Dutch universities and data science centers, research groups exist with the scientific excellence required to address the COMMIT2DATA data science, stewardship and technology challenges. We mention several of the leading research groups and scientists. Several of the above mentioned academic research groups include winners of prestigious VIDI (Lejla Batian/RUN, Birna van Riemsdijk/TUD, Shimon Whiteson/UvA, Joris Mooij/UvA, Andy Zaidman/TUD), VICI/Pioneer (Bart Jacobs/RUN, Maarten de Rijke/UvA, Catholijn Jonker/TUD, Henri Bal/VU, Herbert Bos/VU, Eelco Visser/TUD, Bettina Speckman/TUe) or ERC (Bart Jacobs/ RUN, Herbert Bos/VU, Marieke Huisman/UT, Shimon Whiteson/UvA, Joris Mooij/UvA) research grants, or members of the Royal Netherlands Academy of Arts and Sciences (Jan Bergstra/UvA, Wil van der Aalst/TUe, Inald Lagendijk/TUD).

Further data science excellence exists with Applied Universities (HBOs) and organizations focusing on knowledge transfer. Applied Universities such as HvA and Fontys, have established lectorships that excel in vocational education of big data, as well as accelerating the collaboration in particular with SMEs. To foster the collaboration with industry, so-called Centres of Expertise have been established where applied science activities and valorization projects are conducted in a structural setting. The centers attract industry, large and small, as well as academic scholars that search for realistic settings to validate and valorize there scientific insights. TNO has made the Early Research Program on “Making Sense of Big Data” part of its four years strategic plans (2015-2019). TNO collaborates with many top researchers, is active in the “ICT doorbraakproject Big Data”, and has launched the

Eindhoven University of Technology	• *Process mining* • *Geometry and visualization*	Wil van der Aalst	
		Jarke van Wijk	
UvA	• *Language, image and machine learning* • *Computer vision*	Maarten de Rijke	Max Welling
		Theo Gevers	
VU	• *Knowledge representation* • *High performance computing*	Frank van Harmelen	Henri Bal
		Herbert Bos	
University of Utrecht	• *Games* • *Geometric modeling*	Remco Veltkamp	
CWI	• *Databases*	Stefan Manegold	
University of Twente	• *Interaction technology* • *Data bases* • *Software methodologies*	Vanessa Evers	
		Dirk Heylen	
Delft University of Technology	• *Software engineering* • *Computer systems* • *Machine learning*	Arie van Deurse	Dick Epema
		Marcel Reinders	
University of Leiden	• *Bioinformatics*	Joost Kok	
Radboud University	• *Security* • *Machine Learning*	Bart Jacobs	
		Tom Heskes	
Groningen University	• *Distributed systems* • *Visualization*	Marco Aiello	
		Jos Roerdink	

Big Data Value Center in Almere. The Netherlands e-Science Center (NLeSC) supports public-private collaborative research in data-driven and compute-intensive sciences, with specific focus on life sciences & health, humanities & social sciences, environment and physics.

Dutch academic ICT researchers have a strong record in collaborative project and public-private collaboration. COMMIT is an exemplary research program where 50 private and 20 public partners (including 8 universities) have synergetically collaborated in 16 projects. These projects not only delivered high quality scientific results,; more importantly 50 valorization sprints were executed since early 2014, attracting 30 new private partners to the program. Dutch universities also have a very good track record in creating successful spin-offs. As examples we mention MonetDB and Software Improvement Group (SIG) from CWI, GameMaker (now part of YoYo Games) from University of Utrecht; EUVision (now part of QualComm, and currently in the process of establishing a deep learning QualComm lab in Amsterdam) from University of Amsterdam; Infotron from TU Delft; and SecurityMatters from University of Twente.

Chapter 4

Program Design and Budget

Program Design and Budget

COMMIT2DATA is structured as three interlinked program lines that each contribute to driving Dutch innovation in big data in its own right. The program lines each address different stakeholders with appropriate instruments, from high-science and high-tech industrial research stakeholders pushing for new products and services, to low-tech SMEs who are owners or users of potentially valuable data. The use-inspired research program line is a main driver of the program and addresses the shared data science, stewardship and technology research challenges of Chapter 3 in companies and organizations in the four economic and societal sectors of Chapter 2. The valorization sprint program line brings the solutions created in the first program line, in COMMIT2DATA's precursor COMMIT, and in related data-science research programs to fruition the high-tech and high-science companies in a broad range of economic and societal sectors. Finally, the dissemination program line aims at practical aspects of new and existing big data solutions for (low tech) SMEs. For the valorization sprint and dissemination program lines, an important success factor is the focused and intensive interaction with data owners and businesses active in potentially any top sectors or societal challenge. Though the three program lines address different stakeholders, there will be a constant flow of data science knowledge and expertise from the use-inspired research line to valorization sprints and dissemination. similarly, use-inspiration emerging from dissemination and valorization sprints will be absorbed in use-inspired data science research projects.

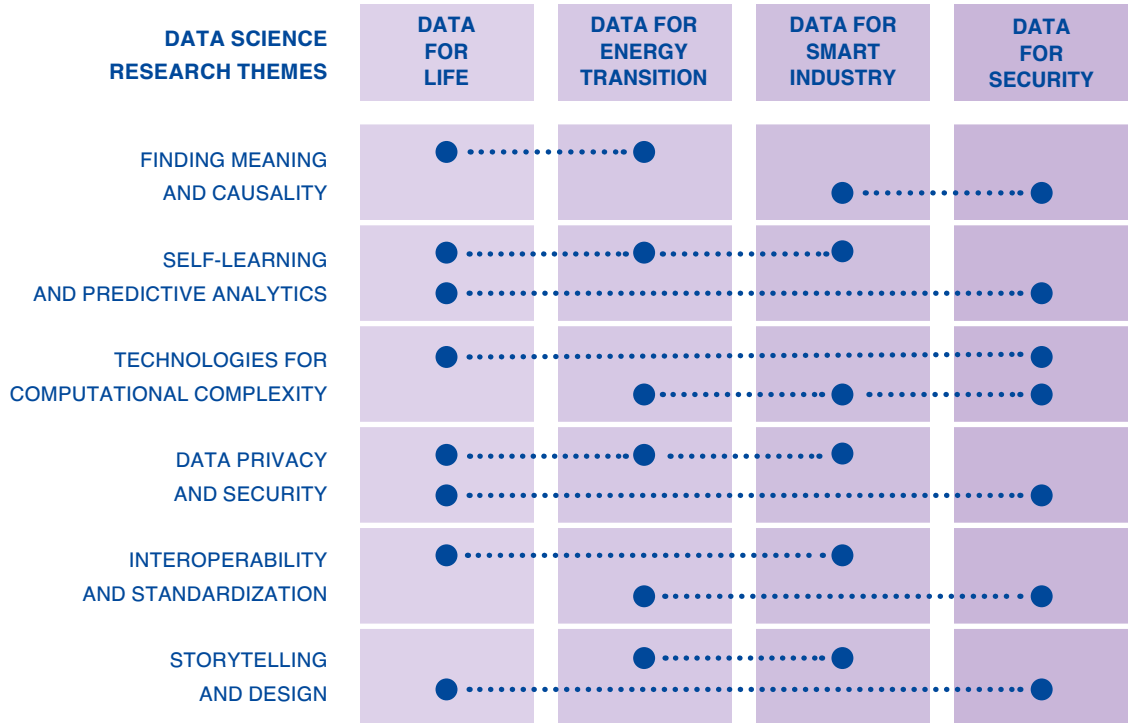
Design of the Use-Inspired Research Program Line

The motto of this program line is: "Develop Knowledge for Big Data". Each of the four sectors of Chapter 2 has specific data science, stewardship and technology challenges for turning big data opportunities into concrete products and services. But these challenges also have a lot of commonalities, see Chapter 3. For that reason, the pre-competitive use-inspired research program line will be addressing data science challenges aiming at multiple sectors at the same time. In this way the scientific level and lateral translation of research results across the sectors will be maximized. At the same time synergy between the sectors is embedded in the program from the very start, bringing together communities that normally would not connect. Yet the shared big data research themes will open up new solution avenues and stimulate valorization across a wide range of economic sectors and companies.

As an example, consider the FAIR concept in life science and health applications. This concept potentially translates well to other sectors that have data trust, access, and privacy challenges, which includes smart industry (see "big data – big trust") and security. At the same time, the demands and earlier solutions per sector can enrich the FAIR concept as so to make it more versatile. Similar arguments will hold for the other data science challenges of Chapter 3.

A following graphical representation of the resulting COMMIT2DATA program structure builds on the two dimensions discussed in Chapter 1: horizontally the context of the use of the big data (application domain), and vertically the properties of the data and the subsequent objectives of data science, stewardship and technology. The horizontally connected bars symbolize COMMIT2DATA projects, addressing one (or more) research themes and addressing multiple sectors at the same time.

GRAPHICAL REPRESENTATION OF DESIGNED COMMIT2DATA PROJECTS IN USE-INSPIRED RESEARCH



This figure is for illustrative purposes only. In the context of this white paper, specific COMMIT2DATA projects have not yet been composed. Once the various public and private funding commitments are firm, a series of sectorial and cross-sectorial open match-making meetings will be organized in order to stimulate and assist the forming of cross-sectorial projects. These projects will then be funded and launched based on a number of proven evaluation criteria – *for instance in the COMMIT program*. These criteria involve review of science quality and uniqueness, IP prospect, potential to contribute to the valorization sprint and dissemination program lines, synergy plans, private funding commitment, balanced distribution of the projects across research themes of Chapter 3 and sectors of Chapter 4, and connection to European big data agenda.

As shown in the COMMIT program repeatedly, research and valorization in economic and societal sectors can benefit highly from creative sector approaches such as the use of social media, gamification and co-creation in design. For that reason, use-inspired projects to be developed will explicitly incorporate researchers and experts from the creative sector to maximize synergy between these sectors and to contribute to the T-shaping of COMMIT2DATA researchers and projects.

Projects in the pre-competitive use-inspired research program line run typically 4 to 5 years and take the form of the round-table or joint strategic programming model (see Knowledge and Innovation Agenda ICT, Chapter 4), involving collaborating public (company researchers and innovation managers) and private (academic researchers, Ph.D. students, and postdoctoral) research-oriented partners.

Design of the Valorization Sprint Program Line

The motto of this program line is: "Transfer Knowledge of Big Data". Valorization sprints are relatively short (6 to 24 months) activities that use cutting-edge science and technology results of COMMIT2DATA projects as a catalyst for focused demonstration and pre-development projects in a wide range of companies and sectors. Since the science of big data is one of the most rapidly progressing fields in recent years, renewal of high-tech and high-science companies is constantly needed using the results of data science research. Valorization sprints are designed to do just that. Valorization sprints are also a means for grounding and cross-fertilization of (existing and new) results of related data-science research programs and the earlier COMMIT research program. This mechanism will be strengthened by explicitly targeting valorization sprint projects on partners of these related data-science research programs.

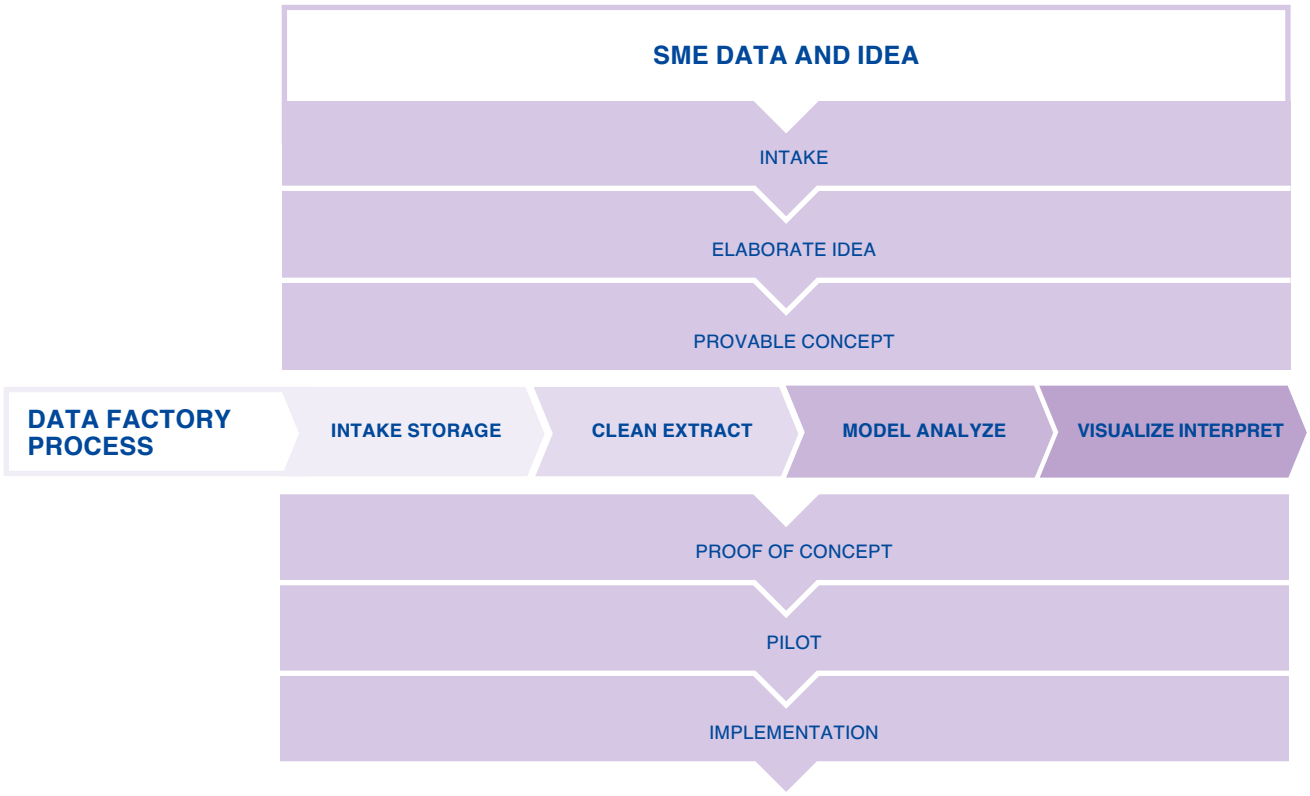
Since valorization sprints aim to implement knowledge spill-over in many sectors, the program line will be highly dynamic. It aims to have a wide range of companies involved, and have series of open calls for proposals. Big data will trigger changes in businesses where new entities like data-brokers and third party service providers will create new (manufacturing) service players. It is essential that new and smaller businesses (start-ups, spin-outs) get involved in these big data breakthroughs and implement them as soon as possible. Existing (manufacturing, OEM, operating) companies need the flexibility of these more agile partners who understand big data research and can adapt it to practical use. The valorization sprints are therefore also a means for acceleration of delivering new high-science and high-technology knowledge or results to market through high-tech SMEs or start-ups.

The proposal submission and evaluation mechanisms will be based on the best practices of the STW "Take off" program, of EIT ICTLabs/ Digital, and of the COMMIT valorization procedure. Essential in valorization sprint is that the take-up of results is by an outside party – *possibly new to the COMMIT2DATA consortium* – that sees new opportunities for the use of results in addition to the ones already part of the use-inspired research program line. For that reason, the evaluation of valorization sprint proposals will be focused on leveraging high-science and high-technology knowledge or results, innovativeness of use, tangibility of "golden demonstrator" or pre-developed product, market and change potential, and transparent planning.

Design of the Dissemination Program Line

The motto of this program line is: "Do Big Data". The dissemination program line aims at maximally efficient spreading of the possibilities that big data offers. The program line builds on the best practices of the "ICT doorbraakproject Big Data", of the COMMIT program (in particular the mid-term event about "The Big Future of Data"), and of other SME-oriented initiatives. An important feature of the dissemination program line is that it targets companies and organizations that are data and problems owners but that do not have high-science and high-tech capabilities themselves. In other words, they need help in creating maximal value out of their data.

BIG DATA FACTORY AS PART OF COMMIT2DATA DISSEMINATION PROGRAM LINE



Whereas traditional dissemination takes the form of broadcasting knowledge through stage presentations and demonstration, the core of the COMMIT2DATA dissemination program line emphasizes hands-on interaction with stakeholders. These stakeholders are typically smaller businesses for which the use-inspired research and valorization sprint program lines are “too academic”. On the one hand they need practical experience with data science results and tools, on the other hand, they also need to work with cutting edge technologies in order for their business to be “big data ready”.

The program line follows three main strategies. The first strategy is that of helping SMEs to gain experience with (their own) big data via “big data factories”. In these factories businesses, knowledge institutes and government interact in developing and executing data experiments. Data factories are regionally oriented to maximize coverage and attractiveness to local SME. They also stimulate mobility between companies and knowledge institutions as so to stimulate the creation of a circular innovation ecosystem. Seeds for COMMIT2DATA big data factories are already available. We mention a number of relevant initiatives.

- Big Data Value Center Almere, involves among others TNO, various big data companies, Windesheim, KvK, SURFsara;
- Data Science Alkmaar, involves among

others city of Alkmaar, companies, VU, innovation hotspot “de Telefooncentrale”;
- The Hague Campus, involves among others city of The Hague, Yes!Delft, VNO-NCWest, innovation hotspot Leiden;
- Big Data Eindhoven, involves among others cities of Eindhoven and Den Bosch, Fontys, ZLTO;
- Dutch Game Garden, involves among others Media Park, KvK, Economic Board Utrecht/ Hilversum, companies, University Utrecht, Hogeschool Utrecht, HKU. iMMOvator;
- Target, involves among others Astron, RUG, city of Groningen;
- Amsterdam Creative Industries Network with nine application labs. Involves among others applied science universities, big and small digital data driven companies such as Digitas LBi, Cisco, Bell Labs Europe and Info.nl.

Big data factories can operate on location, on tour, and on line. These big data factories also play a role in realizing vocational education and training.

The second strategy is “Big helps Small & Small helps Big”. The aim of this sponsorship strategy is to stimulate growth in small big data companies and, simultaneously, big companies that are still small in the use of big data. Through collaborative pilots involving both small and big companies, the small companies will be able to grow faster than currently is the case, and big companies will change faster thanks to the technology provided by the fast moving small companies.

The final strategy involves the organization of frequent COMMIT2DATA national events, contributions to national and international big data event, and thematic workshops bringing together elements of the three program lines.

Connection to Related Initiatives

In the further development of COMMIT2DATA and the composition of concrete projects and activities in the three program lines, alignment with related national and international initiatives will be pursued. In the European context, the Big Data Value Association (BDVA) and the European PPP on Big Data are important. COMMIT2DATA has the ambition to play are role in the shaping of the work program of BDVA, and projects executed as part of the European PPP Big Data. Already several Dutch parties in applications of big data and data science are member of BDVA. In fact, thanks to the choice of the economic and societal top sectors involving big data for energy transition, for smart industry, for life sciences, and for security, public and private organizations participating in COMMIT2DATA will be better positioned for a role in BDVA and PPP Big Data. In order to emphasize the European connection and alignment, part of the COMMIT2DATA budget will have to come from European funding, see COMMIT2DATA budget.

Nationally, the CHAT and ERCET initiatives aim at a broad technological spectrum ranging from astronomy and humanities to life sciences and energy. In terms of application domains, these initiatives are partially complementary and partially in agreement to the COMMIT2DATA sectors. But more importantly, at the core of the CHAT and ERCET initiatives is the need for data science, stewardship and technology research rather similar to the ones outlined in Chapter 3 of this white paper. It is therefore well possible to extend the design of the valorization sprint and/or use-inspired research program lines to intimately connect to data science challenges of CHAT and ERCET. This perspective offers the opportunity to shape a national big data research and innovation program covering economic, societal and scientific value.

COMMIT2DATA Budget

The table below expresses the budget that will be needed for execution of COMMIT2DATA over a period of five years (2016-2020). A growth scenario is foreseen in which initial annual commitments are lower than requested, but which is compensated for in later years as COMMIT2DATA successfully attracts new partners. The funding for the COMMIT2DATA program relies on private (company) funding, regional, national and European governmental funding, and matching first money stream of knowledge organizations (not included in the table). Especially for the pre-competitive use-inspired research program line, the ambition is to have 15% co-funding from European research programs, such as PPP Big Data and EIT Digital. This level of co-funding corresponds to the multiplier factor 5 to 6 that the PPP Big Value and EIT Digital aim for. The private contribution of the use-inspired research program line constitutes 25% of the budget.

The budget for the valorization sprint program line is sufficiently large to be able to accommodate valorization projects associated to other top sector, societal challenges or sciences. As explained in Chapter 3, valorization sprint projects will also be developed in collaboration with partners from the COMMIT precursor program and with related data science programs in order to maximize the fruition of data science in high-tech and high-science companies. The private contribution of the valorization sprint program line constitutes 50% of the budget. For the dissemination program line the private contribution is between 20% and 25% depending on the size of the companies involved in the "data factory" initiatives. Overall, the COMMIT2DATA program aims at 30% private funding.

BUDGET COMMIT2DATA

PROGRAM LINE (total for period 2016-2020)		GOVERNMENT (including NLeSc and CWI) WO	TNO	EUROPE (PPP Big Data; EIT Digital)	REGIONAL CONTRIBUTIONS	PRIVATE CONTRIBUTIONS	Budget (M) - Total	
USE-INSPIRED RESEARCH	15-20 projects year each	41 M	18 M	10 M	15 M	-	25 M	**109 M**
VALORIZATION SPRINTS	20 projects	-	7 M	6 M	-	3 M	15 M	**31 M**
DISSEMINATION	100 SMEs in data factory; events	-	5 M	4 M	-	2 M	3 M	**14 M**
Budget (M) - Total		**41 M**	**30 M**	**20 M**	**15 M**	**5 M**	**43 M**	**154 M**

Appendix

Contributors

Contributions:

Team ICT

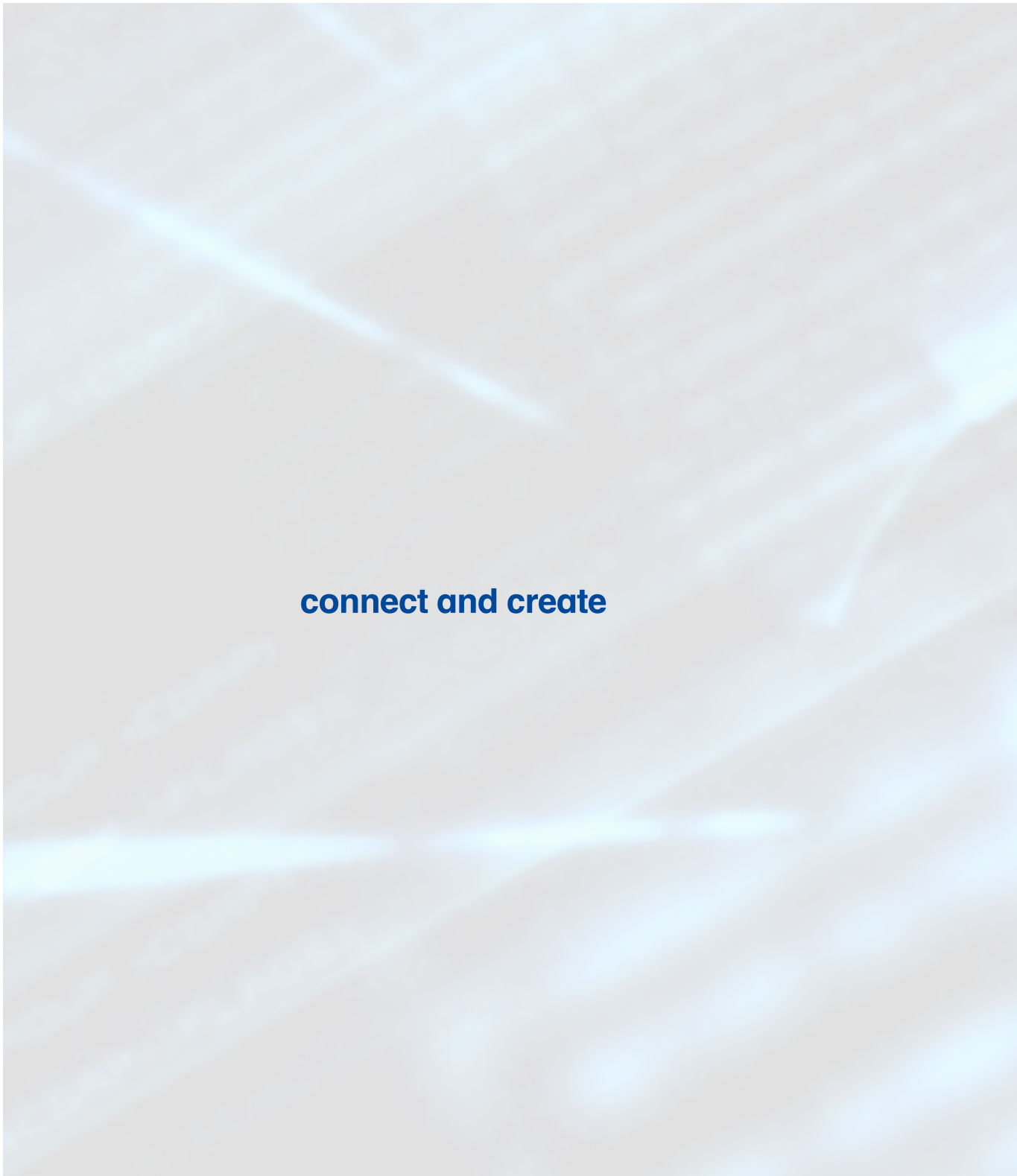
René Penning de Vries	Captain
Inald Lagendijk	Captain of Science, Delft University of Technology
Ineke Dezentjé Hamming-Bluemink	FME-CWM
Ben Woldring	Bencom Group
Gerben Edelijn	Thales Netherlands
Mark Bressers	Ministry of Economic Affairs
Erik Wijnen	Ministry of Economic Affairs
Arie van Bellen	ECP

Ops Team

René Penning de Vries	Captain
Robert van der Drift	NWO
Henk-Jan Vink	TNO
Ronald Verbeek	CIO Platform Nederland
Lotte de Bruijn	Nederland ICT
Arie van Bellen	ECP
Erik Wijnen	Ministry of Economic Affairs
Ben van Lier	Centric

Thanks to

Wil van der Aalst	TU Eindhoven
Peter Apers	Univ. Twente & COMMIT
Helen Burger	Politie
Robert van der Drift	NWO
Bert Feskens	HSD
Robin de Haas	HSD
Robin Hagemans	Alliander
Jaap Heringa	VU Amsterdam
Geert-Jan Houben	TU Delft
Rene Kamphuis	TNO
Ruben Kok	DTLS
Geleyn Meijer	HvA & COMMIT
Barend Mons	Univ. of Leiden
Milan Petkovic	Philips
Han la Poutre	CWI
John Post	TKI Urban Energy
Marcel Reinders	TU Delft
Rene Hooiveld	TNO
Tobias Paulissen	V&J/NCTV/DCS
Gera Pronk	ICT Doorbraakproject Big Data
Maarten de Rijke	UvA
Arnold Smeulders	UvA & COMMIT/
Egbert-Jan Sol	Smart Industry Team
Maarten van Steen	Univ Twente & IPN
Jeroen van der Tang	Nederland ICT
Arie van Tol	TNO
Eric van Tol	ICT Doorbraakproject Big Data
Frits Verheij	DNVGL & TKI Urban Energy
Henk-Jan Vink	TNO
Erik Wijnen	Ministry of Economic Affairs


connect and create

dutch
digital
delta