


Use-Inspired Research on Big Data and Applications in the Public-Private Research and Innovation Program Commit2Data

Boudewijn R. Haverkort ✉ 

Tilburg School of Humanities and Digital Sciences, Tilburg University, The Netherlands

Aldert de Jongste ✉

ECP, The Hague, The Netherlands

Pieter van Kuilenburg ✉

ECP, The Hague, The Netherlands

Abstract

In this paper we give an overview of the public-private research and innovation program known as Commit2Data, which was executed throughout the years 2016 – 2024 in the Netherlands. We outline the set-up of the program, with special attention for its valorisation activities, and provide a future outlook.

2012 ACM Subject Classification Information systems; Applied computing

Keywords and phrases Big data, public-private partnership (PPP)

Digital Object Identifier 10.4230/OASICS.Commit2Data.2024.1

Acknowledgements Running the Commit2Data program has been quite an endeavor, but has also been very fulfilling. We are grateful for the generous financial support from NWO, the Ministry of Economic Affairs and Climate, and the Topsector ICT. Over the years, many dedicated staff members of these organizations, as well as from the Netherlands Organization for Applied Scientific Research (TNO) and ECP supported the Commit2Data program generously with their time and expertise. We are grateful for their efforts, which have been instrumental to the success of the Commit2Data program.

1 Introduction

Data science is an interdisciplinary academic field, focusing on the collection of data sets from systems or processes, extracting information from such data sets, and further developing such information to knowledge that can be applied in solving a wide range of problems [4]. Data science became recognized as an independent academic discipline around the 1990s. Since then, the amount of data generated on a daily basis has increased in a way that is hard to phantom, primarily through the emergence of the internet, in particular the world-wide web [1]. More and more people globally are using the internet on a daily (hourly, minute) basis, and the advent of the internet of things has resulted in additional data generation and data transfer that requires virtually no human interaction; this is nicely visualized in Figure 1. As a result, the total amount of (meta-) data generated around the world on a daily basis is currently estimated at several hundred million terabytes. Since 1998, the collecting, handling and processing of such large amounts of data is often referred to as “big data” [2, 7].

However staggering the amount of data available, data in itself is of little economical and societal value. Only when data is analysed, turned into information, then into knowledge and finally applied to specific problems, it becomes valuable in a wide range of application domains. In order to reach this value, expertise from the application field has to be brought together with expertise on big data handling and big data analytics. It is exactly this



© Boudewijn R. Haverkort, Aldert de Jongste, and Pieter van Kuilenburg;
licensed under Creative Commons License CC-BY 4.0

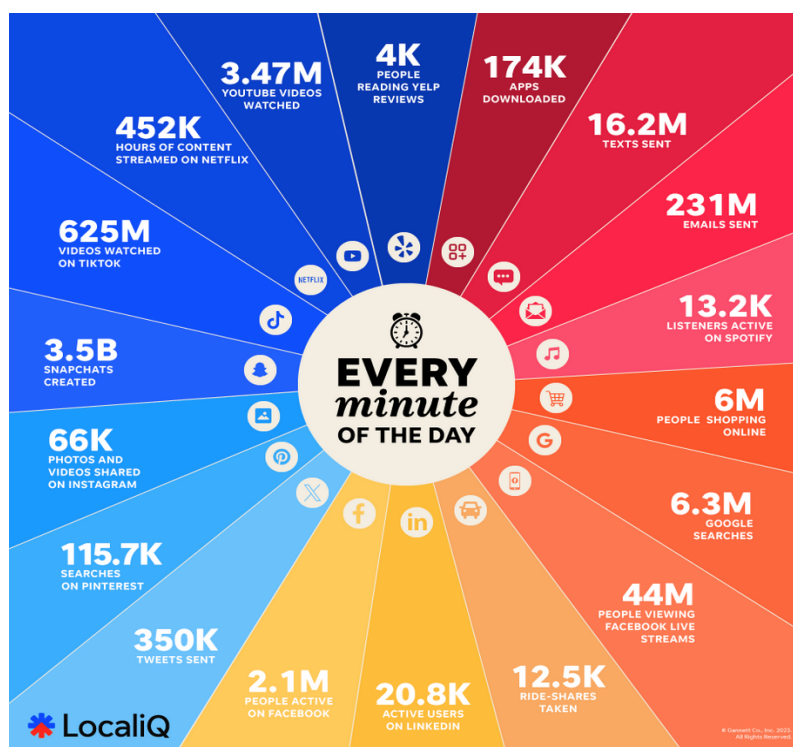
Commit2Data.

Editors: Boudewijn R. Haverkort, Aldert de Jongste, Pieter van Kuilenburg, and Ruben D. Vromans; Article No. 1;
pp. 1:1–1:8



OpenAccess Series in Informatics

OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



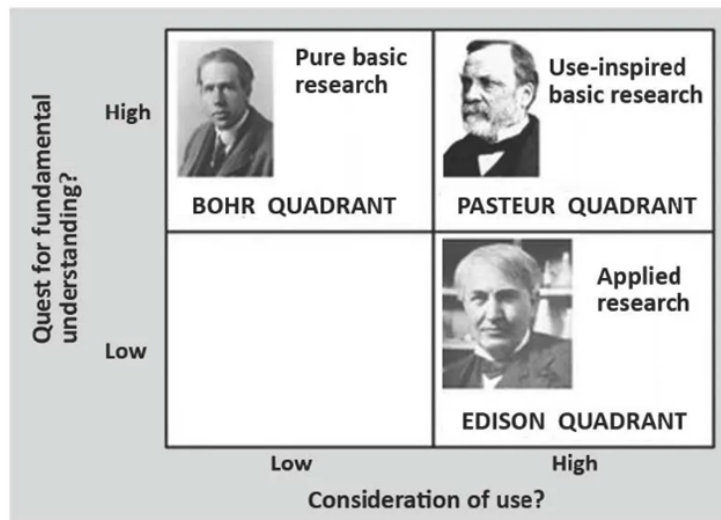
■ **Figure 1** The amount of data generated per minute on the internet in 2024. Figure from <https://localiq.com/>.

combination that was put at the core of the Commit2Data public-private research and innovation program that was set-up throughout 2015 in the Netherlands [8], and that started in 2016.

This paper is further organized as follows. We will present the set-up of the Commit2Data program in more detail in Section 2. Subsequently, in Section 3, we will specifically address the valorisation activities we developed in the program. We conclude the paper with an outlook in Section 4. The papers following this paper will address a selection of projects from the Commit2Data program, across various application domains, in more detail.

2 The structure of the Commit2Data program

The Commit2Data program has been set-up in a joint effort of the Ministry of Economic Affairs and Climate (EZK), the Netherlands Organization for Scientific Research (NWO) and the Dutch national applied research and technology organization (TNO) in 2016. At that time, the expected impact of big data handling and big data analytics for all sectors, not just the ICT sector itself, was deemed enormous. It was also clear that many sectors were taking steps towards further datafication and digitalisation, yet many sectors were also struggling with this transformation. The aim of the Commit2Data program was in that sense twofold: to stimulate fundamental research in big data handling and analytics, as well as to stimulate the uptake of such advanced techniques in a variety of sectors, in particular



■ **Figure 2** The four quadrants as described by Stokes; Commit2Data focuses on “use-inspired basic research”, also known as the Pasteur quadrant.

the so-called Dutch topsectors.¹ By connecting the key enabling technology “big data” to the well-organized topsectors, a win-win situation was created: big data scientists, primarily coming from (applied) computer science departments at Dutch universities could connect to key applicants (often private parties) of such techniques, leading to fruitful cross-fertilization. In doing so, in the terminology of Stokes [6], so-called “use-inspired research” was performed, cf. Figure 2: basic research, yet inspired by concrete use-cases.

The Commit2Data program aimed to bring together (i) excellent data science, with (ii) specific application domain knowledge and inspiration. This is essential because the field of big data is delineated in two dimensions. The first dimension considers data properties and objectives, such as volume, heterogeneity and quality of data. Each of these comes with its own body of knowledge and scientific challenges. Together they define the science of data, data stewardship and data technology in any big data application. The second dimension is the application domain and context in which big data is to be used: contextual information is needed to understand the specific properties and limitations of the data and its intended use. Applications may be very different because of dissimilarities between application domains, yet the underlying subset of data science challenges is often very similar indeed. By addressing multiple application areas, as done in the Commit2data program, cross-fertilization and mutual gains can be obtained.

Figure 3 presents the overall structure of the program. Next to the initial two horizontal subprograms, each focusing on key challenges in the area of big data, without any specific application context in mind, most projects were developed in other subprograms (verticals), organized per sector. Per subprogram (in Figure 3: for every horizontal and for every

¹ The so-called topsectors are economic sectors with high activity and value creation in the Netherlands, responsible for a large share of the Dutch gross national product; see <https://www.topsectoren.nl/> (only in Dutch).

1:4 The Commit2Data Research & Innovation Program

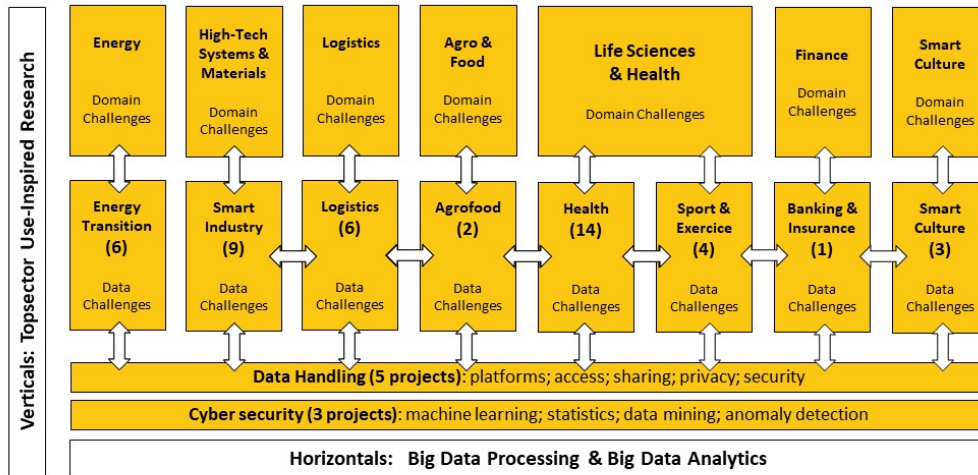
vertical), open calls for research projects were issued. These calls were executed by NWO (the Dutch organisation for scientific research funding), however, the scope of the call was developed by dedicated working groups involving computer scientists as well as scientists from the respective application context. The subsequent call execution, including the selection of projects by NWO through independent reviews and independent selection panel making the final decisions, ensured high standards of scientific integrity. In the figure, the number of executed projects per sector is indicated; a project typically runs for 4–5 years, involves multiple universities and public and private partners, and gives financial room for the appointment of multiple PhD candidates and/or postdocs. The chairperson of the Commit2data program, Boudewijn Haverkort, was not involved in any of the selection procedures, nor did he participate in any of the projects himself, in order to maintain his independent role.

Before every call deadline, open matchmaking events were organised, to bring interested scientists together with scientists and engineers from the application fields. The funding received in each accepted project, was to be used for the appointment of PhD candidates or postdocs at the participating universities. Next to funding from NWO, participating public or private parties from the application domains, had to provide in-cash and in-kind co-funding. The amount of co-funding depended on the sector involved; some sectors are better equipped for high co-funding than others. The overall program budget amounted to 61M€, out of which 18M€ came from public and private partners, hence, we almost attained 30% overall co-funding. In this way, true public-private collaboration was created. This co-funding has multiple effects: for one, there is some influence of the external parties on the topics addressed in the project, however, the mere fact that these parties do really invest, makes their commitment to the project larger, e.g., in making data sets available, giving access to technical experts, etc. For the researchers, the possibility to work with real data, on industrial size cases, also puts their methods to the test, much more than in a smaller scale laboratory setting. Private parties involved in new technologies, like big data, also appreciate their interaction with universities, in order to keep abreast of developments and to position themselves as interesting employer. Finally, the involvement of the private parties from the outset of the project definition, makes that the research projects have high application potential; this has also been witnessed in the valorisation program, as will be discussed in Section 3.

Next to the organisation of research and valorisation, considerable effort was put on dissemination at program level. The aim has been to make the research conducted in the Commit2Data program accessible to a wider audience than just the consortium partners and readers of scientific articles. The dissemination output of the program encompasses a website² with extensive project updates, interviews with project leaders and researchers, various publications in trade journals, from healthcare to logistics to agriculture, and several animations that positions a project in a societal context and illustrate the practical societal value of academic research for a broader public.

Two parents: success factor and challenge. As stated, a key objective of the program was to organize use-inspired research. To achieve this, consistent efforts were made to connect data science knowledge and expertise, with in-depth knowledge of the application areas. To accomplish this, for most of the open calls, an exploration was conducted by a representative from the application domain, roughly addressing the question “What are the most significant

² The Commit2Data projects can all be accessed at <https://commit2data.nl/en/>.



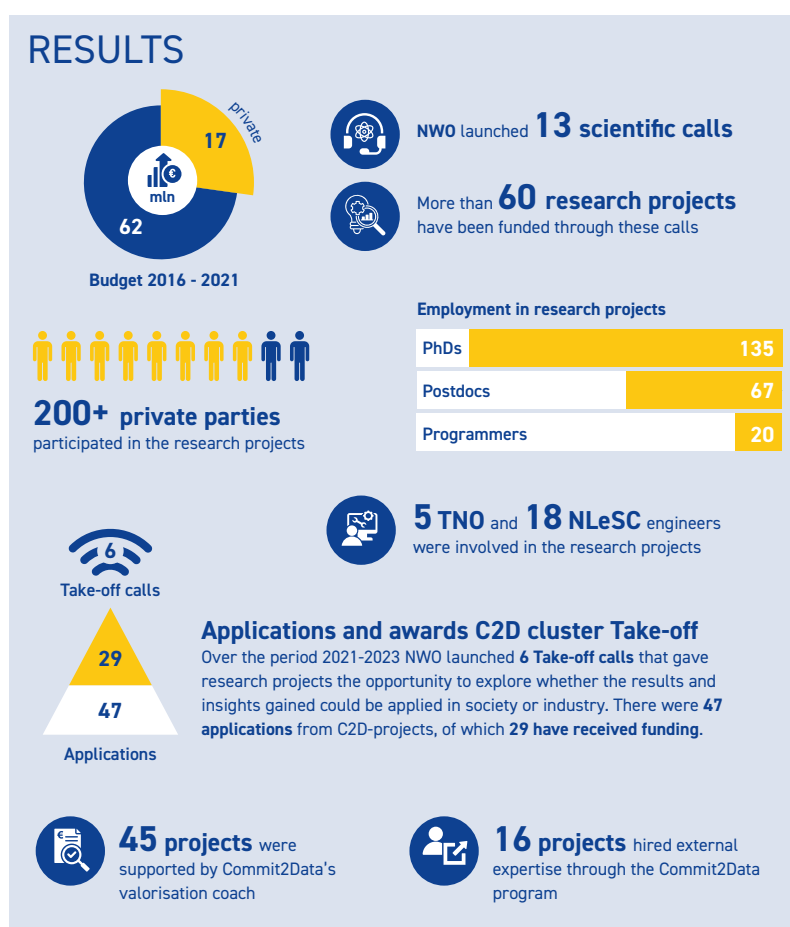
■ **Figure 3** Overall structure of the Commit2Data program, with each block representing a number of thematically connected projects.

data-related questions within this domain?” Based on the outcome of such an exploration, we brought together sector-specific funding (e.g., from the topsector Energy, and the topsector Logistics) and Commit2Data funding. Together, they collaboratively drafted a call text, brought it to the attention of relevant parties, and organized a matchmaking event.

Whereas the two-parent approach has been instrumental in achieving the program’s focus on true applications, it also has had a drawback. Many of the projects and many of the researchers in the projects, identify themselves as sector-specific projects (or researchers). They are primarily focused on healthcare or logistics (to name two examples), because that is where they feel their main network and interests lie. A resulting disadvantage of this is the perceived lack of cohesion within the Commit2Data research and innovation program as a whole. Originally, the idea was that knowledge exchange between sectors could also be organized, meaning that some sectors could benefit from results achieved in other sectors. After all, the idea is that a data analytics solution in healthcare might also work well for a similar problem in smart industry. However, the implementation of this cross-fertilization turned out to be challenging in practice. Project members felt their connection along the sector-specific line rather than along the data analytics line, and knowledge sharing outside the sectoral ecosystem turned out to be difficult to organize. As coordinators of the overall program, we would have liked to see this differently, and we still think that opportunities are missed, however, we did not manage to change this throughout the program. To conclude this section, Figure 4 visualizes some key indicators of the overall program.

3 Valorisation activities

Knowledge valorization is the process of creating value from knowledge, by making knowledge suitable and/or available for economic and/or societal use, by translating it into easy-to-use products, services, processes and new business [9]. Although the valorisation of research results was one of the program’s objectives from the outset, no specific budgets were initially allocated for it. We therefore developed valorisation activities as a separate component from 2019 onwards, with separate funding. That was around the time that the initial projects started to deliver results suitable for valorization.



■ **Figure 4** Visual of the overall output of the Commit2data program.

The development of the valorisation program raised several questions, e.g., about target groups, instruments and financing. For example, whom should we support, just the consortia as a whole or also individual project participants, like researchers, private companies or other societal actors? This in turn led to questions about the financing structure; none of the existing instruments in the Netherlands catered to all these different groups. Finally, a robust valorisation program could be established, building on the NWO Take-Off program, but the late start did have some drawbacks. Due to the late start, project consortia could only be informed and involved at a late stage of the project execution. This meant that the consortia could only incorporate new possibilities into their approach in this later stage, and plan their staffing accordingly. Researchers are often committed for four years only; with full prior knowledge of the possibilities of the valorisation program, they can better plan ahead. A lesson learned from the Commit2Data program is therefore to have a clear understanding of what the follow-up to the research phase will be, already at the start of a program or call, so that project consortia can factor the use of such valorization options into their approach.

Although funding was pledged in various stages, all partners agreed on a valorisation plan by the end of 2019, which received an extra financial impulse from the ministry of Economic Affairs and Climate (close to 2 M€). The original plan included a variety of instruments, from raising awareness and influencing perceptions to customized support for projects or

individuals facing specific challenges. The more generic instruments, such as valorisation workshops, were found to have limited reach within the project consortia, and by the year 2000, we transitioned fully to a customized approach. The central figure in this approach became the so-called valorisation coach. The valorisation coach, which we hired specifically for this purpose, approached project groups, starting with those projects that would come to an end first and were therefore closest to the potential useful results. Projects that started later were approached later. This approach was also partly based on the availability of capacity for support.

The valorisation coach had various tools for support, including his own experience and network. In practice, he often assists researchers, but also startups or consortia in sharpening their valorisable results and making valorisation challenges more manageable. The plan included additional financing so that topic-experts could be hired for specific questions to help and to advise projects or even develop proof of concept implementations. On several occasions, an experienced entrepreneur was sought and found who could coach a startup with the help of our funding.

Substantial effort was put into setting up a valorisation granting scheme, through which researchers could apply for up to 40 K€ to develop a research idea further into a product or first prototype for demonstration purposes, as a first step to further societal or commercial exploitation. A special call for the purpose was issued twice per year in the period 2022–2024 (in total 6 calls) via the NWO Take-Off program. From within the Commit2data program, 46 applications were submitted for such support, of which 27 were granted. The Commit2Data valorisation coach did play an active role in encouraging and supporting projects to submit good proposals to these calls. As a result, not only is the number of applications exceptionally large, also the success rate is very high, which we explain by the fact that our projects have been set-up to be use-inspired, and by the active involvement of our valorisation coach, respectively. Figure 4 also summarizes the key output of the valorization program.

4 Future outlook

The field of data science has undergone significant evolution, marked by the proliferation of big data and the increasing prominence of artificial intelligence (AI). In contemporary discourse, many of the techniques developed in the context of big data research, are now subsumed under the broader and more fashionable term “AI.” This conflation, although technically inaccurate describes the current state of practice.³ Many of the initiatives that started in the Commit2Data program, have found their way to the Netherlands AI Coalition (NLAIC),⁴ and their funding schemes, like AINed⁵.

Even though the term “big data” is less fashionable these days, big data is poised to continue its impact across various sectors. Advances in data storage and processing technologies, such as cloud and edge computing, are expected to handle even larger and more complex datasets with greater efficiency. Further integration of big data with technologies like the internet-of-everything⁶ and the wider availability of bandwidth for mobile communication

³ In the interview [5], machine-learning pioneer Michael Jordan explains why today’s artificial-intelligence systems are not actually intelligent; see also [3].

⁴ See <https://nlaic.com/en/>; the NLAIC is a public-private partnership in which knowledge institutes, government agencies, societal organisations and commercial companies work together to accelerate AI developments in the Netherlands and connecting AI initiatives.

⁵ See <http://ained.nl/en/about-ained/>.

⁶ See <https://ioe.org/>.

will enable real-time data collection and analysis on an increasing scale. As the interweaving of big data technologies with our everyday lives grows, so does the need for proper ethical considerations. To safeguard data privacy and security, innovative and robust frameworks will have to be developed to ensure responsible data governance. As these trends unfold, the distinction between big data and AI will likely blur further, fostering innovative applications and insights.

The Commit2Data program played a significant role in advancing methodologies and applications with big data. The approach developed through Commit2Data are now more commonplace in a range of funding calls by the Netherlands Organisation for Scientific Research (NWO). These calls reflect a broader integration of data science principles, emphasizing the importance of data-driven decision-making and the application of sophisticated analytical techniques across diverse scientific fields. This progression underscores the dynamic nature of data science and its foundational impact on contemporary research paradigms.

References

- 1 Tim Berners-Lee, Robert Cailliau, cois Groff Jean-Fran and Bernd Pollermann. World-Wide Web: The Information Universe. *Internet Research*, 2(1):52–58, 1992. doi:10.1108/eb047254.
- 2 Francis X. Diebold. On the origin(s) and development of the term “Big Data”. *Penn Institute for Economic Research Working Paper*, 12(37), 2012. doi:10.2139/ssrn.2152421.
- 3 Michael I. Jordan. Artificial Intelligence – The Revolution Hasn’t Happened Yet. *Harvard Data Science Review*, 1(1), July 2019. <https://hdsr.mitpress.mit.edu/pub/wot7mkc1>. doi:10.1162/99608f92.f06c6e61.
- 4 Chantal D. Larose and Daniel T. Larose. *Data Science Using Python and R*. Wiley, 2019.
- 5 Kathy Pretz. Stop calling everything AI. *IEEE Spectrum*, March 2018.
- 6 Donald E. Stokes. *Pasteur’s Quadrant – Basic Science and Technological Innovation*. Brookings Institution Press, 1997.
- 7 Michael Stonebraker, David Blei, Daphne Koller, and Vipin Kumar. ACM Panels in Print: Big Data. *Communications of the ACM*, 60(6):24–25, 2017.
- 8 Topsector ICT. Commit2data white paper: Proposal for a national public-private research and innovation program on data science, stewardship and technology across top sectors. Technical report, Topsector ICT, the Netherlands, 2015. URL: <https://ecp.nl/publicatie/commit2data-white-paper/>.
- 9 Leonie van Drooge and Stefan de Jong. Valorisation: onderzoekers doen al veel meer dan ze denken. Technical report, Rathenau Instituut, Den Haag, 2015. URL: <https://www.rathenau.nl/nl/kennis-voor-transities/valorisatie-onderzoekers-doen-al-veel-meer-dan-ze-denken>.