Late to the Party?

Agile Methods in British and German Government Institutions

Master's Thesis

in fulfillment of the requirements
for the degree

Master of Public Policy

at the Hertie School, Berlin

Author:

Maximilian Kupi

Matr. No. 183587

Supervisor:

Prof. Slava Jankin

Mikhaylov, PhD

Professor of Data Science and
Public Policy

Berlin, 25th May 2020

Executive Summary

This thesis project developed an online data collection and analysis system to trace the evolution and spread of agile methods in British and German government institutions. Agile methods are rooted in the software development departments of the private sector and are characterized by an iterative development process that focuses on the user. As the digital transformation gains ground in the public sector, these methods are also becoming more relevant to government institutions. Yet, public management literature still lacks an understanding of the temporal and spatial evolution of this development. To fill this research gap, this study analysed the occurrence of agile methods related keywords on British and German government websites over time. A total of 49 government domains were crawled and 171,569 potentially agile related pages downloaded. After preprocessing (e.g. extracting the publishing organisation's name and the publishing date) and cleaning the data, 451 relevant sites were left. The analysis showed that the number of agile related sites published by government institutions as well as the number of government domains publishing agile sites increased over time and particularly in recent years (289% increase in number of published sites from 2017 to 2019). Furthermore, it revealed that 84% of agile sites also mention digital transformation related keywords. German federal and state level institutions published a total of 74 agile related pages, the first one in 2015, and the most active bodies turned out to be the Federal Ministry of Labour & Social Affairs and the state of Baden-Wuerttemberg. The Federal Government Office and the Bavarian government lead in terms of most detailed methods description. On the other hand, Britons have been discussing agile methods on their government web presences since 2011 (380 pages in total), while the lead institutions in terms of number of sites published are the Government Digital Service and the Cabinet Office. With regard to the depth of agile methods descriptions the leaders are the Home Office and the Education & Skills Funding Agency. When comparing British and German central government ministerial departments, the wide lead – in terms of all three, quantity, timing, and depth – of the British institutions becomes apparent. To catch up, Germans should more strongly acknowledge the strategic relevance of agile methods for government's digital transformation, and consider establishing distinct lighthouse institutions that internally push the application of agile methods and openly talk about it so that other institutions can follow. Future work could include further relevant actors such as NGOs, consultancies or media outlets in the analysis to possibly reveal who "drives" the adoption of agile methods in government institutions.

Contents

List of Figures	iv
List of Tables	v
Acknowledgements	vi
1 Introduction	1
2 Literature Review	3
2.1 Digital Transformation	3
2.2 Agile Methods	4
2.3 Benefits, Challenges, & Conditions	8
2.4 Research Gap	9
3 Methods	11
3.1 Study Design	11
3.2 Data Collection	11
3.3 Data Preprocessing	13
3.4 Data Analysis & Visualisation	17
3.5 Prototyping & Testing	18
4 Analysis	20
4.1 General Level	20
4.2 Intra-Country Comparison	20
4.2.1 Germany	20
4.2.2 United Kingdom	23
4.3 Inter-Country Comparison	27
5 Discussion	30
5.1 Summary of Findings	30
5.2 Policy Implications	32
5.3 Limitations & Future Work	33
6 Conclusion	35
References	vii
Appendix A	xiii
Appendix B	xv
Appendix C	xxii
Statement of Authorship	xxiv

List of Figures

1	Waterfall development process	4
2	Agile development process	6
3	Percentage of agil* sites referring to digital transformation	20
4	Number of Agil* sites published and publishing domains over time	20
5	Agil* sites published by German state and federal government institutions over time	21
6	Agil* sites published by German federal government institutions over time	21
7	Average number of co-search terms per agil* site published by German federal government institutions	22
8	Breadth and depth of German state government institutions' web publica- tions referring to agile methods	22
9	Agil* sites published by British government institutions over time	24
10	Agil* sites published by British central government ministerial departments over time	25
11	Average number of co-search terms per agil* site published by British gov- ernment institutions	26
12	Agil* sites published by British and German central government ministerial departments over time	27
13	Breadth and depth of agil* sites published by British and German central government ministerial departments	28
14	Frequency of co-search term lemmas for all agil* sites published by British and German central government ministerial departments	29
15	Breadth and depth of agil* sites published by British and German central government ministerial departments (with institution names)	xiii
16	Agil* sites published by British central government ministerial departments over time (excluding sites from gov.uk's national web archive)	xiii
17	Agil* sites published by British and German central government ministerial departments over time (including sites from gov.uk's national web archive)	xiv
18	Breadth and depth of agil* sites published by British and German central government ministerial departments (with institution names and including sites from gov.uk's national web archive)	xiv
19	Frequency of co-search term lemmas for all agil* sites published by British and German central government ministerial departments depicted as word clouds	xiv
20	Scrapy Architecture	xxii

List of Tables

1	Values of agile software development	5
2	Principles of agile software development	7
3	Conditions for applying agile methods in public institutions	9
4	Summary of findings	30
5	List of government websites	xv
6	List of English and German co-search terms	xviii
7	Number of sites per state / country before and after preprocessing	xxiii

Acknowledgements

I would like to express my heartfelt appreciation to my supervisor, Prof. Slava Jankin Mikhaylov, PhD, who supported me in word and deed during the entire thesis project. Also, I warmly thank my colloquium peers Sören Borghardt, Huy Dang, Anna Dietrich, Allison Koh, Yuan Ting Lee, Alexandra Magaard, Maximilian Rekuts, and Ole Teutloff for their valuable feedback as well as for the interesting and entertaining discussions during our sessions. Furthermore, I wish to give special thanks to Hari Krishna Vetharenian for consulting me on python programming and web scraping, and Henriette Kupi for proofreading the manuscript. Finally, I would like to express my extraordinary thanks to Ines Rieger for her academic as well as emotional support.

We will always be looking to improve, because the expectations of citizens and the demands on services will always be shifting.

Kevin Cunningham, Director General,
Government Digital Service, UK (2017)

1 Introduction

The digital transformation is continuously changing more and more aspects of contemporary life. From the way people communicate to the way they find the right path or even the love of their life – nothing remains untouched by digital technology. In line with this, digital interaction channels have seen an increase in relevance for the private and public sector alike (Macaya et al., 2019).

In order to digitally transform any organisation, a simple transfer of analog practices into the digital realm will not do: Much rather digital transformation requires organisational and relational changes in order to better satisfy the users' needs and fully reap the benefits that digital technologies can bring (Mergel et al., 2019). In that respect, the adoption of agile methods has increasingly become the mantra for many public sector institutions (Lundqvist & Olesen, 2016; Mergel, Whitford et al., n.d.; Vacari & Prikladnicki, 2015).

Agile approaches originate from the IT departments of private sector companies, and are characterised by the incremental and iterative development, testing, and improvement of technology products with a focus on the user (Mergel, 2016). However, as more and more non-IT related teams have come to value digital technologies as integral part of their work, agile methods such as Scrum or Kanban began to expand beyond these departments. In that respect various digital government units such as United Kingdom's, United States' or Canada's government digital service are regarded as the lead organisations that "paved the way" for other government departments to also introduce agile methods (Mergel, Whitford et al., n.d., p. 2, see also Clarke, 2019). However, apart from these prominent examples, canonical public administration literature knows little about the extent to which these methods are actually prevalent in other government institutions (Carvalho Fernandes et al., 2016; Mergel et al., 2018; Mergel, Whitford et al., n.d.; Vacari & Prikladnicki, 2015). To help close this research gap, this study will analyse the appearance of agile related

keywords on British and German government websites over time and answer the following research question:

RQ: *What is the evolution and spread of agile methods in British and German government institutions?*¹

With its much cited digital transformation efforts, the United Kingdom will thereby serve as the baseline against which German government institutions will be assessed (Clarke, 2019; Sivarajah et al., 2014). To provide the most legitimate basis for comparison, only central government ministerial departments' web sites will be analysed for the inter-country comparison. However, in order to gain further insights, subordinate British institutions and German state governments will also be included in the respective intra-country comparisons.

The remainder of this paper is structured as follows: The literature review provides an introduction to public sector's digital transformation as well as to the respective British and German policy setting. Furthermore, the characteristics of agile methods as well as the connected implications for their application are discussed, and the research gap is specified. The study design as well as the data collection, preprocessing, and analysis methods to answer the research question are described in Section 3. Additionally, this section outlines the prototyping and testing work that went into the development of this analysis pipeline. Section 4 presents the analysis results on a general level as well as for the intra- and inter-country comparisons. The discussion section summarizes the findings, derives policy implications, and points out alleys for future research based on this studies limitations and beyond. The paper ends with a short conclusion.

¹When talking about the United Kingdom or Britain, this paper refers to all the government institutions that have their web presences on gov.uk, although, strictly speaking, Scotland (gov.scot), Wales (gov.wales), and Ireland (gov.ie) are also all part of the United Kingdom and have their own government institutions respectively.

2 Literature Review

2.1 Digital Transformation

The term digital transformation has become ever more prevalent in the private and lately also in the public sector to describe the ongoing changes enabled by the adoption of digital technologies (see for example Behörden Spiegel, 2020; McKinsey, 2018; Tabrizi et al., 2019). Based on interviews with 40 experts from the field, Mergel, Edelmann, and Hauga (2019, p. 12) define the digital transformation of public administrations as "a *holistic* effort to revise core processes and services of government beyond the traditional digitisation efforts" [emphasis added]. To further specify the term, they differentiate digital transformation from the two connected concepts of digitisation and digitalisation (Mergel et al., 2019; see also Bloomberg, 2018; Brennen and Kreiss, 2016).

While digitisation depicts the one-to-one transition of analogue services into the digital realm, digitalisation goes further as it focuses "on potential changes in the processes beyond mere digitising of existing processes and forms" (Mergel et al., 2019, p. 12). Digital transformation, in contrast, encompasses both these elements, but puts a strong emphasis on organisational and relational changes of the digitally transforming entity in order to better satisfy the user needs (Mergel et al., 2019; see also Henriette et al., 2015; Vial, 2019). In line with that, Mergel, Edelmann, and Hauga (2019) identify the change in culture, skills, and mindset as an essential condition to make the digital transformation of public administrations last (see also Dierks, 2020; Parcell and Holden, 2013). With their strong focus on staying responsive to change and satisfying user needs efficient- and effectively, adopting agile methods (see Section 2.2) is considered key (Andriole, 2018; Beyer, 2010; Mergel, 2019; Mergel et al., 2018; Osmundsen, 2020).

At the European level the "Tallinn Declaration on eGovernment" (2017b) "marks a new political commitment [...] towards ensuring high quality, user-centric digital public services for citizens [...]" (European Commission, 2017a). Recognising the need to "increase the digital leadership skills" and "modernise the design of public services, procurement and contracting arrangements, to make them compatible with modern and agile ways of developing and deploying digital technology" on the country-level (European Commission, 2017b, p. 7), the declaration can be seen as the latest supranational push towards the digital transformation of public administration (Mergel et al., 2019). Additionally, the European Union passed the Single Digital Gateway Regulation in order "to provide access

to information, to procedures and to assistance and problem-solving services” for citizens and businesses in all EU member states (European Union, 2018).

In line with these mandates, Germany released the Act for the Improvement of Online Access to Administration Services (Onlinezugangsgesetz), which requires all public services to be available digitally until 2022 (Deutscher Bundestag, 2017). Addressing nearly 600 public services to be digitalised in a user-centred manner, this law can be considered the main driver for public administration’s digital transformation in Germany up until now (Bundesministerium des Inneren für Bau und Heimat, 2017; Mergel, 2019; see also European Commission, 2019a).

However, compared to the UK, current efforts in Germany focus mostly on technology and place less emphasis on a real digital transformation of the public administration (European Commission, 2019b; Mergel, 2019). Already bearing ”transformation” in its title, the vision of UK’s Government Transformation Strategy is to ”transform the relationship between citizens and the state – putting more power in the hands of citizens and being more responsive to their needs” (Cabinet Office of the United Kingdom, 2017). In order to achieve this, the strategy puts a strong emphasis on growing ”the right people, skills and culture”, and clearly acknowledges the importance of agile methods (Cabinet Office of the United Kingdom, 2017).

2.2 Agile Methods²

Traditionally, the development of IT related projects followed the linear waterfall approach: Each phase in the development process (see Figure 1) is tackled sequentially, one at a time (Kannan et al., 2014; Sherrell, 2013). The core assumption behind this approach is that by eliminating any possible mistakes within each phase, the subsequent phases ”won’t be impacted by mistakes and the project team won’t lose time and money by going back to fix the mistakes” (Mergel, 2016, p. 517). As a consequence of this rigid structure, the waterfall process is unresponsive to sudden changes in

Figure 1: Waterfall development process (adapted from Mergel, 2016)

²This section has been adapted from a passage the author previously submitted for assessment in a paper on ”Applying Agile Development Methods in E-Government Projects” for Professor Dr. Mehmet Akif Demircioglu at the Lee Kuan Yew School of Public Policy, Singapore.

needs and requirements (Kannan et al., 2014). Moreover, evaluations of the project status through user testing of the product or service only happen close to the final launch in the later stages of the process. Only then, the project team might come to realize major flaws, which in some cases consequently leads to a complete failure as in the case of the US' HealthCare.gov project (Mergel, 2016). All these properties make the waterfall approach rather unsuited for large projects, whose nature rarely turns out to be truly linear and sequential in reality (Kannan et al., 2014).

Agile development approaches, in contrast, involve creating, testing, and improving technology products incrementally (Mergel, 2016). The first example of such an approach dates back to the 1920s, when physicist and statistician Walter Shewhart of the Bell Labs developed the Plan Do Study Act (PDSA) to iteratively improve products and processes by responding to changes and new findings throughout the whole development process (Goldman et al., 1994). Currently used agile methods include Scrum, Kanban, eXtreme programming (XP), Lean Software Development, and Feature-Driven Development (FDD) (Dingsøyr et al., 2012; Rigby et al., 2016b). Design Thinking is often seen as a useful complement to further foster the user-centredness of the solution – although, strictly speaking, the method goes beyond the core of agile development methods (Australian Digital Transformation Agency, 2019; Da Silva et al., 2011).

Table 1: Values of agile software development

<i>Value 1</i>	Individuals and interactions over processes and tools
<i>Value 2</i>	Working software over comprehensive documentation
<i>Value 3</i>	Customer collaboration over contract negotiation
<i>Value 4</i>	Responding to change over following a plan

Source: Agile Manifesto, 2001.

By and large, all agile approaches adhere to the tenets of the Agile Manifesto (2001), which was formulated by a group of 17 thought-leaders in 2001 (Rigby et al., 2016b). Table 1 depicts the core values of this manifesto (Dingsøyr et al., 2012). The first value highlights the importance of the people in the development team as well as their continuous interaction and communication. Valuing this over a pedantic obsession with processes and tools is supposed to make the team more responsive to sudden changes. Value number two emphasizes the adoption of a lean mentality when it comes to minimizing unnecessary work, in particular with regards to wasteful documentation procedures. This does not mean eliminating the documentation completely, but streamlining it in a way that provides the developer what is needed to do the work without getting caught up with unnecessary details (Eby, 2016). The third value stresses the importance of collaborating with users

and stakeholders vis-à-vis formal contract negotiations. This way it is supposedly ensured that their needs and concerns are truly understood and addressed by the development team. Finally, the fourth value acknowledges that uncertainties are a constitutive part of any software development project, and thus the team should stay flexible throughout the process to make changes.

To achieve this high level of responsiveness to change, the agile development process is made up of continuous iterations, so called "sprint cycles" (see Figure 2). At the beginning of an agile project the team conducts qualitative field research – potentially using

Figure 2: Agile development process (author's visualisation)

Design Thinking methods – to gather the needs of the users and document them as so called "user stories" in the language of the users in order to avoid misconception and translation errors (Mergel, 2016; Wirdemann et al., 2017). In each of the sprint cycles the project team then develops increments of working software that solve these user needs and are the pieces of the puzzle that ultimately make up the final service or product (Schwaber & Sutherland, 2020). Each of these sprints usually takes between one to four weeks and, in principle, follows the phases of the traditional waterfall approach, however, with a strong focus on iteratively testing the increment (Hughes, 2013; Schwaber and Sutherland, 2020). Small prototypes are used to rapidly (dis-)prove ideas before many resources are invested in their further development, following the motto of failing fast, early and often (Boehmer et al., 2017; Mergel, 2016). The initial project plan serves as a rough guidance. Yet, the early evaluations reveal which requirements might not have been completely fulfilled and need to be added as work packages to the next sprint cycles, leading to a revision of the plan after each iteration (Mergel, 2016). In consequence, the short sprint cycles paired with the continuous testing of the product or service are supposed to reduce the monetary risk involved in software projects (Schwaber & Sutherland, 2020).

In a typical agile project, the development team is self-organising in an environment that enhances creativity and spontaneity. According to Takeuchi, one of the pioneering scholars of agile methods, and Nonaka (1986), three conditions have to be met in order to foster self-organising teams similar to the ones in start-ups: First, *autonomy* as to how exactly the team plans to meet the user needs and requirements with the service or

product. Second, *self-transcendence* by establishing their own goals and amending them continuously based on user feedback as the service or product takes shape. Third, *cross-fertilization* achieved through an interdisciplinary, cross-agency, and cross-functional team constellation that unites members with different skill-sets, thought process, and behaviour patterns. In addition to these three conditions the management has to exercise "subtle control" in order to enhance a team's creativity and spontaneity. This means establishing "enough checkpoints to prevent instability, ambiguity, and tension from turning into chaos", while at the same time avoiding the kind of rigid control that hinders the flow of the creative juices (Takeuchi and Nonaka, 1986). In most agile projects, the team's self-organisation is institutionalized by regular project meetings, as for example the daily stand-ups or the sprint planning and review meetings at the beginning and end of each sprint cycle in the case of the Scrum methodology (Scrum.org, 2019). The 12 principles of the Agile Manifesto (2001) represented in Table 2 summarize the aforementioned particularities of agile methods in a comprehensive manner.³

Table 2: Principles of agile software development

<i>Principle 1</i>	Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
<i>Principle 2</i>	Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
<i>Principle 3</i>	Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
<i>Principle 4</i>	Business people and developers must work together daily throughout the project.
<i>Principle 5</i>	Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
<i>Principle 6</i>	The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
<i>Principle 7</i>	Working software is the primary measure of progress.
<i>Principle 8</i>	Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
<i>Principle 9</i>	Continuous attention to technical excellence and good design enhances agility.
<i>Principle 10</i>	Simplicity – the art of maximizing the amount of work not done – is essential.
<i>Principle 11</i>	The best architectures, requirements, and designs emerge from self-organising teams.
<i>Principle 12</i>	At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behaviour accordingly.

Source: Agile Manifesto, 2001.

³The words "customer" and "business people" in the wording of the principles hint at the methodology's origin in the private sector domain. For a public sector set-up "customer" might be translated to "recipient of the government service or product" and "business people" might be translated to "public manager".

2.3 Benefits, Challenges, & Conditions

It has to be recognized that – besides all the potential benefits the approach can bring – going agile is definitely not a panacea under all circumstances (Rigby et al., 2016a). Furthermore, the approach can even pose certain challenges to an organisation, in particular when it comes to applying it in government institutions whose values, structures, and processes might oftentimes be especially hard to combine with the agile spirit (Hajjdiab & Taleb, 2011).

On the one hand, literature has identified various potential *benefits* that agile approaches can bring to government institutions. The most mentioned ones are as follows: First and foremost, scholars have found that agile methods lead to a better and more collaborative integration of IT teams with public sector management (Berger, 2007; Berger et al., 2005; Dubinsky et al., 2005; Upender, 2005).⁴ Second, agile methods are found to increase the satisfaction of government services' recipients and other important stakeholders (Fruhling et al., 2008; Fulgham et al., 2011; Iliev et al., 2009). Third, researchers have observed that an agile approach in government projects reduces the time required for value delivery (McMahon, 2006; Moore, 2001; Surdu & Parsons, 2006; Upender, 2005). Fourth and finally, agile approaches were found to improve team morale and increase public servants' job satisfaction (Dubinsky et al., 2005; Hajjdiab & Taleb, 2011; Vacari & Prikladnicki, 2015).

On the other hand, research shows that applying agile approaches in public sector institutions can pose a number of *challenges*: Most important in that respect is the often perceived cultural mismatch between the rigidity of bureaucratic line organisations and the flexibility of agile methods (Altukhova et al., 2016; Berger et al., 2005; Fruhling et al., 2008; Greve et al., 2019; Mergel, Whitford et al., n.d.). Second, since agile methods still constitute a novelty in public sector domains, the lack of education and experience in their application is another common challenge identified by scholars (Fridman, 2016; McMahon, 2006; Nuottila et al., 2016). Third, the shortage of (senior) management support can also pose difficulties to agile projects in the public sector (Berger, 2007; Dubinsky et al., 2005; Hajjdiab & Taleb, 2011; Mergel, Whitford et al., n.d.). Fourth and finally, the necessary adaption of procurement and contracting arrangements as well as the compliance of the agile working mode with relevant standards and regulations can be a difficult task for public sector institutions that intend to go agile (Fruhling et al., 2008; Fulgham et al.,

⁴As Sørensen and Torfing (2011) have presented in their well received article, collaboration in turn can lead to an enhancement of public sector (digital) innovativeness.

2011; Mergel, Whitford et al., n.d.; Nuottila et al., 2016).

Ultimately, it is important to acknowledge which are the *conditions* that are most likely favourable for applying agile methods, and under which circumstances the methodology is unlikely to yield benefits. In their Harvard Business Review article "Embracing Agile", Rigby, Sutherland, and Takeuchi (2016a) have stated various factors that are important in that respect. Adapted for public sector organisations, these factors are summarized in Table 3.

Table 3: Conditions for applying agile methods in public institutions

<i>Conditions</i>	Favourable	Unfavourable
<i>Environment</i>	User preferences and solution options change frequently.	Conditions are stable and predictable.
<i>User Involvement</i>	Close collaboration and rapid feedback are feasible. Users know better what they want as the process progresses.	Requirements are clear at the outset and will remain stable. Users are unavailable for constant collaboration.
<i>Innovation Type</i>	Problems are complex, solutions are unknown, and the scope is not clearly defined. Product specifications may change. Creative breakthroughs and delivery time are important. Cross-functional collaboration is vital.	Similar work has been done before, and innovators believe the solutions are clear. Detailed specifications and work plans can be forecast with confidence and should be adhered to. Problems can be solved sequentially in functional silos.
<i>Modularity of Work</i>	Incremental developments have value, and users can utilise them. Work can be broken into parts and conducted in rapid, iterative cycles. Late changes are manageable.	Users cannot start testing parts of the product until everything is complete. Late changes are expensive or impossible.
<i>Impact of Interim Mistakes</i>	They provide valuable learning.	They may be catastrophic.

Source: Adapted from Rigby et al., 2016a.

2.4 Research Gap

Due to its growing relevance for practitioners, the application of agile methods in the public sector has increasingly become the focus of researchers around the world. However, recent literature reviews suggest that these studies are still limited in number and scope – most of the scholars employ case studies, yet, few also make use of surveys or interviews to derive their insights (Carvalho Fernandes et al., 2016; Mergel et al., 2018; Mergel, Whitford et al., n.d.; Vacari and Prikladnicki, 2015). Furthermore, while the existing studies uncover

useful insights on the micro / organisational level, macro level studies that quantitatively analyse the evolution and spread of agile methods in government institutions are still lacking. This paper aims to help closing this research gap on the basis of an intra- and inter- country level comparison for the United Kingdom and Germany.

(This white space is to illustrate how big the research gap actually is.)

3 Methods

3.1 Study Design

In recent years, online text mining has become an important field in social science research (Ignatow and Mihalcea, 2018, pp. 37 ff. Shumate and Weber, 2016). The application of the technology spans from social media knowledge extraction (A. Salloum, 2017), to analysing leadership behaviour in e-learning contexts (Xie et al., 2018) or conducting systematic literature reviews (Mergel et al., 2018).

Building on Michael Thelwall's idea of "Webometrics" (Thelwall, 2009), this study employs an exploratory, primarily quantitative research design based on the analysis of self-crawled web content (see also Ignatow and Mihalcea, 2018; Jaeger and Halliday, 1998; Olston and Najork, 2010).⁵ To answer the research question on the evolution and spread of agile methods in German and British government institutions, the websites of the respective institutions were crawled and analysed with respect to the appearance of the term "agil*"⁶ (following the keyword search strategy of Mergel et al., 2018). Thereby gathered insights, in particular the number of agile related sites published as well as the detail in which the methodology is being described on these sites, are meant to serve as a proxy measure for the evolution and spread of agile methods in the respective institutions (Branco and Rodrigues, 2006; Ghosh et al., 2013). The following three sub-sections describe the methods applied for the pipeline of data collection, preprocessing, and analysis – all of which were programmed and deployed by the author in the programming language Python (Van Rossum and Drake Jr, 1995).⁷ The final prototyping subsection briefly outlines all the testing and prototyping work that went into developing this pipeline.

3.2 Data Collection

The purpose of the data collection process was to efficiently gather as many potentially relevant websites as possible. The process followed the steps identified for routine web crawls by Schäfer and Bildhauer (2012). As the first step of data collection, a list of all official websites of federal and state cabinets and ministries in the UK and Germany was gathered manually by consulting the respective web presence listings from official

⁵The reason for not relying on crawling results from search engines such as Google is that these results are deliberately biased and thus significantly impair the reliability of the results (Cothey, 2004).

⁶The star is used as a placeholder for all potential English and German endings of the word, e.g. agile, agility, agiles, Agilität etc.

⁷The code is publicly accessible on <https://github.com/MaximilianKupi/agile-in-government>.

government websites (see Table 5 in Appendix B). These websites then were crawled with the help of the open source web crawling platform *Scrapy* (Kouzis-Loukas, 2016), specifying the main URLs from which the state / federal government institutions can be reached (e.g. <https://www.gov.uk> or <https://www.bayern.de> etc.) as the "seed URL" (Barbaresi, 2015, p. 115). These seed URLs serve as the starting point for the crawler: All further pages linked on this page will be crawled subsequently and so on and so forth until a previously specified depth limit is reached or the links leave a previously specified domain range. In order to exclude non-governmental web sources, the crawl was limited to stay within the range of the respective domain of the seed URL as well as all its sub domains.⁸ For efficiency purposes and building on the assumption that institutions would not hide relevant information on deep pages of their websites, links with greater depth than ten pages were ignored (Scrapy, 2018; Wang et al., 2019). Furthermore, the crawler was set to ignore the `robots.txt` file, since for many sites effective web crawling would not have been possible otherwise (Sun et al., 2007; see also Barbaresi, 2015, p. 125). The general architecture of Scrapy is outlined in Appendix C. To search for the respective keyword on all crawled pages as well as to retrieve and store the relevant web content, the spider was setup as follows:

1. The crawled site was downloaded as HTML file and parsed into a "soup" object with the help of the *Beautiful Soup* package (Richardson, 2007).
2. The text of the HTML file was extracted.
3. Based on the CONDITION if the search term agil*⁹ was present or not, the following steps were performed:
 - (a) IF the search term was present:
 - i. A random ID for the page was created based on its URL and stored as an item for later processing in the pipeline together with the URL and domain name.
 - ii. Four different (potential) publishing date instances, as well as the document's main heading were extracted with the help of XPath expressions and stored as items for later processing in the pipeline (w3schools.com, 2020).
 - iii. The HTML file was stored in a newly created directory under the newly created ID.

⁸E.g. <https://www.gov.uk> and <https://gds.blog.gov.uk/> etc. or <https://www.bayern.de> and <https://km.bayern.de> etc.

⁹In regex: `\bagil.*` .

iv. The path name was reported as "saved file" in the log file.

(b) IF the search term was NOT present:

i. The page was being skipped (reported as "skipping" in the log file).

4. The process repeated from step 1 with the next crawled site.

To not get stuck trying to decipher non-HTML files, all other common file extensions (e.g. `.flv`, `.css`, `.pdf`) were denied for extraction. After the HTML files were parsed and analysed in the spider, the items were passed to the items pipeline (see Scrapy architecture in Figure 20 in Appendix C) which wrote them into a CSV file and stored the file for further processing.

The final crawls were run on a specifically set up instance on the Google Cloud Platform from servers in Belgium, and closely monitored throughout the whole process via an https connected JupyterLab. File syncing with the local machine was implemented via rsync (Tridgell & Mackerras, 2020). All in all the crawler downloaded 171,569 pages.¹⁰

3.3 Data Preprocessing

The purpose of the data preprocessing was to strip the text from unnecessary web-script elements, extract further date timestamps while selecting the earliest one, and finally, delete duplicates and false positives from the dataset, and save everything in machine readable CSV- and text-file formats. These steps are in line with the recommendations of Lüdeling et al. (2015, p. 19) for the preprocessing of downloaded web pages. In order to achieve maximally reliable results, a mix of automated and manual preprocessing tasks was employed. The automated methods were self implemented and adapted for their language / country specific application. The following two paragraphs list and describe the methods in detail.

Automated Preprocessing Tasks:

- Deleting cases with corrupted HTML files. In order to do so, the encoding format was detected using the *chardet* package (Pilgrim et al., 2015). In case parsing the file with the respective encoding still yielded a decoding error, the file was deleted from the dataset.¹¹
- Retrieving further instances for (potential) release date entries from the web pages so that all relevant cases would be equipped with at least one date entry. To achieve

¹⁰For the numbers per state / country see Table 7 in Appendix C.

¹¹In total, only one case had to be deleted due to a corrupted HTML file.

this, the parser package *lxml* (Faassen, 2006) was used besides *Beautiful Soup* in order to be able to read XPath (Clark and DeRose, 1999). In addition to the custom coded functions for date extraction on this particular set of websites, the *HTMLdate* package (Barbaresi, 2020) was used in order to extract further standard date instances (e.g. dates in URLs etc.). To correctly parse the dates into Python datetime objects, the *dateutil* package (Niemeyer, 2003) was used.

- Selecting the earliest retrieved date instance as a the finally considered publishing date. To ensure the validity of the assumption that the earliest date instance would match the actual publishing date for the majority of cases, random samples from all relevant sites were drawn and manually checked.
- Removing all HTML code and other "boilerplate" elements such as headers or links so that only the text content would be left for further text analyses (2015, p. 19). This step was performed with the help of the *Beautiful Soup* package (Richardson, 2007), and the so cleaned text was stored as a variable as well as in separate text-files.
- Searching for more specific *agil** terms in the cleaned text with the help of more restrictive regex expressions, and saving the results as a variable.¹²
- Extracting the preceding and subsequent context (four words each) of the first appearance of the search term and saving them as variables in order to facilitate the later search for false positives. This was also done using regex expressions.¹³
- Searching for the appearance of digital transformation related keywords in the texts and storing the results into a variable.¹⁴
- Deleting false positives, i.e. cases that have been downloaded by the crawler but in fact turned out to have nothing to do with agile methods.¹⁵ To do so the following sub-tasks were performed:¹⁶
 - Erasing all cases, where the more specific and restrictive regex match for the *agil** terms (see above) did not yield any results.
 - Searching for co-search terms (e.g. "scrum", "sprint", "design thinking"; see

¹²The regex expression to catch all English matches was specified to be `\bagil\b|\bagility\b`, ignoring case. The German regex expression was specified to `\bagility\b|\bagil\w{0,2}\b|\bagilität\b`, ignoring case.

¹³The regex expression to capture four words before and four words after the first appearance of *agil** was set to `\s*([^\s]+?)\s+([^\s]+?)\s+([^\s]+?)\s+([^\s]+?)\s+agil.*?\s+([^\s]+)\s+([^\s]+?)\s+([^\s]+?)\s+([^\s]+?)\s+`

¹⁴The regex expressions were specified to `\bdigital\b|\bditailization\b|\bdigitalisation\b|\bdigital transformation\b` for English and `\bdigital\w{0,2}\b|\bdigitalisierung\b|\bdigitale [n]? transformation\b` for German, both ignoring case.

¹⁵Exemplary cases are "agile" military in the UK context or "agile Senioren" (Eng: agile seniors) in the case of German websites.

¹⁶For all tasks involving the deletion of cases the *pandas* package (McKinney, 2010) was used for dataframe handling.

Table 6 in Appendix A for the complete list) in the cleaned texts of the sites and deleting cases where no co-search term was found.¹⁷ This dictionary-based approach has proven useful for various text classification and mining tasks (see for example Liu and Zhang (2012) and Tang et al. (2013)). To compile the list of co-search terms, the glossaries from Kusay-Merkle (2018) and Peterjohann (2020, building on Schwaber and Sutherland (2020)) were used as a starting point. Co-search terms were then deleted or added manually, based on experimental exploration of the respective results for the given set of web pages.¹⁸

- Deleting all sites that were no "proper" content pages like search result page or table-previews. To do so pages with the respective URL paths, e.g. `csv/preview` or `/search/`, were erased.
- Erasing cases where the `agil*` term appeared only on parts of the page that were not properly removed by the boilerplate cleaning (see above) – e.g. where the term was mentioned in a side bar drop down menu or in a live twitter embedding. These cases were identified and deleted making use of the `agil*` term context (e.g. post agile context "development (33) agile" in the case of a drop down menu entry).
- Deleting cases from general (not public sector related) job listings as well as from supplier listings based on the respective URL paths, e.g. `/jobs/` or `/supplier/`.
- Erasing German cases that originated from non-German versions of the website – e.g. French or English translations. This was done to ensure that no duplicates would be caused by having the same page in different languages. To do so the cases with the respective URL paths, e.g. `/FR/` or `/breg-en/`, were deleted from the dataset.
- Deleting duplicates based on the heading of the site or the exact matches for the `agil*` terms as well as the context of the first `agil*` term.¹⁹
- Adding the publishers organisation name for each site based on the respective sub-domain name using self compiled dictionaries.²⁰ Furthermore, organisation names

¹⁷In order to ensure that all inflected forms of a co-search term would be matched, the search terms as well as the cleaned text was lemmatized beforehand using the *spaCy* package (Honnibal & Johnson, 2015). This was particularly important for German language to ensure that "agiles Team", "agilen Teams", "agiler Teams" etc. would all be matched. Furthermore, all agile mentioning co-search terms were specified in three different versions (without quotes, with single quotes, and with double quotes) so that 'agile' methods or "agile" methods would also be matched. Finally, the German co-search term list also included all English terms, as it is common to use English instead of German words when talking about agile methods.

¹⁸All in all, more than 100 potentially wrongly erased web sites have been systematically analysed, and the respective co-search terms added to the list.

¹⁹All duplicates were erased so that the site with the oldest publication date was preserved.

²⁰For Britain a list of domain names from gov.uk (<https://assets.publishing.service.gov.uk/government/>

for publications on the general domain "gov.uk" had to be identified using a set of seven CSS and XPath selection patterns to scrape the names from the different website set-ups with the help of the *lxml* (Faassen, 2006) and *Beautiful Soup* package (Richardson, 2007).²¹ Additionally, the scraped names had to be cleaned and adapted in order to match the names already existent in the dictionary, and organisation names for sites where the selectors did not yield any results were added manually.

- Deleting all cases published in 2020. Since the crawl was done in the first quarter of 2020, this was done to ensure comparability over all (fully represented) years.
- Storing the preprocessed datasets into CSV-files, and merging all files into one final master data frame to run the analyses with the help of the *pandas* package (McKinney, 2010).

Manual Preprocessing:

- Deleting false positives that have not been caught by above automated methods. For example, cases where only one agil* term was found and the matched co-search term appeared in a totally unrelated section of a page or cases where an otherwise reliable co-search term was used in a non-agile related context. To do so, all pages with only one agil* term or co-search term match as well as pages originating from domains usually not related to agile methods were analysed manually by inspecting the agil* term context and/or the respective web content online. The so identified false positives were then deleted from the dataset.
- Erasing duplicates that have not been caught by above automated methods. For example, cases where preview/summary pages of blogs incorporated other (also in the dataset included) blog articles or cases where multiple, now outdated versions of the same guide have been preserved in an online archive. For that purpose, all pages with the same context for the first appearance of the agil* term but different total agil* term matches or different headings (so that the automated method could not identify them) were analysed manually by inspecting the respective web content online. The so identified duplicates were then deleted from the dataset.
- Re-specifying wrongly specified publishing dates. To do so, all cases where the different date instances diverged substantially (i.e. dates from different years) were

uploads/system/uploads/attachment_data/file/842955/List_of_gov.uk_domain_names_as_at_28_Oct_2019.csv/preview) was used and missing domains were amended manually on the basis of online research. For Germany there was no such list available and thus all organisation names had to be researched manually.

²¹In case a site was authored by multiple institutions, the first mentioned institution was selected.

analysed by inspecting the respective web content online. The so identified correct publishing date was then saved as the final publishing date variable for the respective case in the dataset.

After preprocessing, a total of 451 pages was left in the dataset.²²

3.4 Data Analysis & Visualisation

The purpose of the data analysis and visualisation was to uncover all relevant patterns in the data in order to answer the paper's research question, and visualise these as well as further supplementary, topic related insights in an intuitive and reader-centred way. The major tasks to achieve this were:²³

- Calculating the percentage of agil* sites also mentioning digitalisation / digital transformation and plotting the results as a pie plot (see Figure 3 in the Analysis Section).
- Analysing and plotting the number of sites published as well as the number of publishing domains / organisations over time (see Figure 4 in the Analysis Section).
- Comparing the publications of all British as well as German institutions over time and visualising the results as stacked bar plots (see Figures 5, 9, 6, 10 in the Analysis Section).
- Analysing the average number of co-search terms for British as well as German institutions and visualising them as point plots (see Figures 7, 11 the Analysis Section).
- Calculating the number of agil* sites published as well as the average number of co-search terms per site for all German state governments, and visualising these as choropleth map (see Figure 8 in the Analysis Section). In addition to *matplotlib* (Hunter, 2007) the *geopandas* (GeoPandas Developers, 2013) and the *descartes* (Gillies, 2020) packages have been used for producing the map visualisations.
- Comparing the number of agil* sites published as well as the average number of co-search terms per site for all British and German central government ministerial departments and visualising the results as scatter plots (see Figures 13 and 15 in the Analysis Section and in Appendix A).
- Calculating the frequencies of co-search terms for all publications of German as well as British central government ministerial departments, and visualising these as word clouds, where a bigger word size depicts a higher frequency (see Figure 19

²²For the numbers per state / country see Table 7 in Appendix C.

²³For all data wrangling tasks the *pandas* package (McKinney, 2010) was used. Furthermore, for all plotting the *matplotlib* (Hunter, 2007) and/or the *seaborn* package (Waskom, 2012) have been employed. Additional packages used are specified in the text.

in Appendix A). For the word cloud visualisation the *wordcloud* (Mueller, 2020) package was used in addition to *matplotlib* (Hunter, 2007).

- Selecting the most diverse color palettes possible for plots with a large number of qualitative distinctions so that the reader could still differentiate the colours (see for example Figure 9 in the Analysis Section). This was done using the *RColorBrewer* package (Neuwirth, 2020) in the programming language R since no comparable python package was available.

All preprocessing, analysis, and visualisation tasks were run on a MacBook Pro 2018 with 2.9 GHz Intel Core i9 Processor and 32 GB of RAM memory.

3.5 Prototyping & Testing

The design of the data pipeline and its methods was an iterative process involving plenty of prototyping and testing work. The following items provide a brief overview:

- Selecting and setting up the most efficient and effective crawler package in order to be able to crawl the huge amount of web pages. At first, test crawls were run in R with the *RCrawler* package (Khalil & Fakir, 2017). However, it turned out that this package was way too slow and inflexible for the task at hand. For this reason, the final crawler was set up using the *Scrapy* package (Scrapy, 2018) in python.
- Performing various crawler test runs on the author's local machine to debug and improve the code before executing the final crawls on Google Cloud Platform.
- Designing the most efficient way to get rid of false positives in the data. At first, a negative (deselecting non-relevant cases), purely manual approach was used. However, due to the amount and diversity of text data this turned out to be too time intensive and error prone. Hence, the positive (selecting relevant cases), dictionary-based approach was developed.
- Amending the code to fetch instances of (potential) publishing dates and finally selecting the date that most likely would constitute the "real" publishing date. As the website buildup differed from institution to institution and from country to country, new parsing modes would have to be added as soon as the old ones did not yield any results for newly crawled pages.²⁴
- Rerunning crawls and/or merging crawls from various runs for domains where the crawler yielded errors. In cases where the crawl did not run properly – e.g. because

²⁴This is also the reason, why some initial dates were parsed during the crawl while others were parsed only at the preprocessing step.

it got stuck on certain page sections with thousands of irrelevant page entries – the crawler had to be restarted and/or the crawling results had to be merged with the results from previous runs.

4 Analysis

4.1 General Level

On a general, cross-national level, it is to be noted, that the vast majority of crawler yielded sites mentioning agile methods (in the following referred to as agil* sites), also mention digital transformation related keywords (see Figure 3). This is in line with the apparent connection between agile methods and the digital transformation of the public sector, as already discussed in Section 2.1.

Figure 3: Percentage of agil* sites referring to digital transformation

Furthermore, as Figure 4 depicts, the number of crawled sites mentioning agile methods published on German and British government web presences has seen an increasing trend over the years, with a particular steep growth since 2017 (289 %). Also, the number of different government (sub-)domains publishing agil* sites has been increasing over time.

Figure 4: Number of Agil* sites published and publishing domains over time

Both these developments can potentially be interpreted as hinting at the rising relevance of these methods for an increasing number of government organisations in Germany and Britain, and thus can be seen as a further confirmation of the themes discussed in current topic related public management literature (see Section 2).

4.2 Intra-Country Comparison

4.2.1 Germany

The crawler yielded a total of 71 agil* sites published by German state and federal ministries.²⁵ As becomes apparent in Figure 5 (see next page), German government ministries, have only started publishing agile related content in 2015.²⁶ In addition,

²⁵For the numbers per state / ministry see Table 7 in Appendix C.

²⁶All stacked and vertical bar or point plots have the total number of agil* sites published by the respective institutions / the total number of pages on which a co-search term appears in brackets after the institution name / term.

Figure 5: Agil* sites published by German state and federal government institutions over time

ture field "innovations in the world of work".²⁷ Analysing the crawler yielded sites in Germany makes apparent that the majority of them discuss or mention agile in a similar manner – as something that is (still) seen as mostly "external" to public administration and predominantly relevant for the private sector only. Due to this one article, the Federal Government Office also ranks highest amongst the federal ministries in terms of average number of co-search terms (e.g. "scrum", "sprint", "design thinking"; see Table 6 in Appendix B) with an average of 9 co-search terms per site (see Figure 7 on next page).²⁸ The number of co-search terms might be read as a rough proxy of how in-depth the agile methodology is being described on the respective site.

When comparing the content published

²⁷Link: <https://www.bundesregierung.de/breg-de/themen/forschung/agil-arbeiten-bei-der-software-entwicklung-276792>.

²⁸However, this average number might be skewed, since the crawler only yielded one article published by the Federal Government Office.

it can be seen that federal level ministries – and amongst them the Federal Ministry of Labour and Social Affairs (13 sites, see Figure 6) – have been the most active in that respect (23 sites in total), and the first ones to publish agile related content in 2015. The first site published is an article of the Federal Government Office (bundesregierung.de) introducing agile as a new working method in software development. Following this line of thought, the article sees agile as a topic to which research related to Germany's "High Tech Strategy 2025" needs to pay close attention to, since these methods play an important role in the fu-

Figure 6: Agil* sites published by German federal government institutions over time

by the governmental web domains of the federal states (see Figure 8), it can be seen that Baden-Wuerttemberg's ministries have published the highest number of websites refer-

Figure 7: Average number of co-search terms per agil* site published by German federal government institutions

followed by Berlin (8), Bremen, and Rhineland-Palatinate (both 6). However, in terms of how many co-search terms appear on average on those sites, Bavaria leads the board with an average of 3.2 co-search terms per site, followed by Hesse (2.75), and Brandenburg (2). In Bavaria's case the high average number of co-search terms is due to two articles of the "Elite Network of Bavaria" – an initiative of the Bavarian State Ministry of Science and the Arts – which describe the methodology in length as part of their educational offers.²⁹ Again, these articles are not related to agile methods in public administration per se. Hesse's high score, however, is due to two pages mentioning the methodology in

Figure 8: Breadth and depth of German state government institutions' web publications referring to agile methods

the context of introducing an innovation laboratory for staff of the regional council Kassel as well as one page mentioning agile methods with respect to the digital transformation

²⁹Links: <https://www.elitenetzwerk.bayern.de/elitestudiengaenge/aktuelles-esg/artikel0/scrum-workshop-bei-hubert-burda-media-433/?L=248>, and <https://www.elitenetzwerk.bayern.de/elitestudiengaenge/aktuelles-esg/artikel0/agile-transformation-412/>.

of the public administration.³⁰ Hence, a direct link between the methods and government institutions is established in Hesse's case. The scraper did not yield any matching pages for Saarland, North Rhine-Westphalia, Saxony, and Lower Saxony.

Thus, as a final, summarising remark on Germany's intra-country comparison, it has to be stated that there are substantial differences between state governments and between central government departments when it comes to the number of agil* sites published as well as the average number of co-search terms appearing on these sites. In recent years, the Federal Ministry of Labour and Social Affairs, as well as the states of Baden-Wuerttemberg, Berlin, Bavaria, and Hesse have turned out to be most active in these respects (see Figures 6 and 7). The average differences between central and federal state institutions, however, are rather minor, hinting at a somewhat federalistic approach to bringing agile into government.

4.2.2 United Kingdom

Looking at the matching sites the scraper yielded from gov.uk – all in all a total of 380 pages³¹ – reveals the following patterns over time (see Figure 9 on next page).³² The very first content on agile methods with the fitting title "Opening up the conversation" was published in 2011 by Directgov's development team. It is the start page of innovate.direct.gov.uk, a former innovation platform of direct.gov.uk – UK's official website back then – to share "ideas for digital development [...], propose ideas for new projects and keep up to date with our [the development team's] latest work".³³³⁴ Interestingly, the second publication was about how Britain's new ICT strategy with a focus on adopting agile methods will help "cutting public spending on Government information and communication technology (ICT) by millions of pounds" – shedding light on a potential motivation for the British government to go agile.³⁵ In line with that, other publications at the beginning of 2011 are the ICT Strategy itself as well as announcements of a change of

³⁰Links: <https://rp-kassel.hessen.de/forumzukunft>, <https://rp-kassel.hessen.de/%C3%BCber-uns/stabsstelle-forum-zukunft/inlab/innovationslabor>, and <https://hzd.hessen.de/leistungen/digitalisierung-erfolgreich-gestalten>.

³¹142 of these pages (37.37%) originate from gov.uk's national web archive, where outdated and deleted web sites are preserved.

³²Although the initial goal of the paper was to only include highest government bodies in the analysis (i.e. cabinet office and ministries), crawling gov.uk as the single government domain for Britain also yielded results of subordinate authorities. In order not to discard this data, and potentially also to uncover relevant insights on the topic, these sites were included in the intra-country comparison of Britain.

³³Link: <https://webarchive.nationalarchives.gov.uk/20110410092517/http://innovate.direct.gov.uk/>.

³⁴To a certain extent, this development team could be viewed as a predecessor to Government Digital Service.

³⁵Link: <https://webarchive.nationalarchives.gov.uk/20110406074635/http://www.cabinetoffice.gov.uk/news/government-save-millions-ict>.

personnel and the plan to develop "a single government domain" (gov.uk).³⁶ Besides Directgov and the Cabinet Office, the only other institution publishing agile methods related content in 2011 was Government Digital Service. This "centre of excellence" was founded in 2011 by the Cabinet Office and commissioned to take the leading role in the British government's digital transformation (Government Digital Service, 2020). Consequently, their first crawler yielded publications have been about them wanting to hire "world-class digital talent", moving into their new office or introducing specific tools applied for agile development.³⁷ It is to be noted that both, the Cabinet Office and Government Digital

Figure 9: Agil* sites published by British government institutions over time³⁸

Service, are the only institutions on gov.uk consistently publishing agile related content

³⁶Links: <https://webarchive.nationalarchives.gov.uk/20110406081839/http://www.cabinetoffice.gov.uk/content/government-ict-strategy>, <https://webarchive.nationalarchives.gov.uk/20110803134807/http://www.cabinetoffice.gov.uk/news/joe-harley-cbe-appointed-chief-information-officer-uk-government>, and <https://webarchive.nationalarchives.gov.uk/20110406074527/http://digitalengagement.cabinetoffice.gov.uk/blog/2011/03/29/towards-a-single-government-domain/>.

³⁷Links: <https://webarchive.nationalarchives.gov.uk/20120404161123/http://digital.cabinetoffice.gov.uk/2011/12/08/new-home-for-gds/>, <https://webarchive.nationalarchives.gov.uk/20120404152514/http://digital.cabinetoffice.gov.uk/2011/10/25/the-second-lever/>, and <https://webarchive.nationalarchives.gov.uk/20121102172608/http://digital.cabinetoffice.gov.uk/2011/09/19/introducing-the-needotron-working-out-the-shape-of-the-product/>.

³⁸Grouping of central government ministerial departments according to <https://www.gov.uk/government/organisations>.

every year since 2011 as well as the ones with the highest number of sites published (a total of 55 and 143 sites respectively, see also Figure 10).

Over time, the crawled sites show a similar pattern as in the case of Germany: The number of organisations publishing content on agile methods generally increases. However, non-central government agencies in Britain are (still) significantly less active in that respect – the councils of Durham, Essex, Leicester, and North Yorkshire as well as the Greater London Authority, and the Local Government Authority for example published only one site each. Also, it seems like the election cycle might be somewhat correlated with the number of published sites: One year after the general elections (in 2015 and 2017) there is a notable increase in sites published (see Figure 9 on previous page).³⁹ The Agile Delivery Community clearly was the leading institution responsible for the spike in 2016. It is a community of practice to inform about, share experiences related to, and discuss agile ways of working, and hence might be seen as a potential vehicle to roll out the methodology to other government institutions.⁴⁰ In line with that assumption, the number of different organisations publishing agil* sites shows a significant growth pattern since 2016.

Figure 10: Agil* sites published by British central government ministerial departments over time

Also, it is noteworthy that 42 percent of the agile content was published on online

³⁹The first increase in 2013 could potentially also be linked to the general elections in 2010, however the effects are only noticeable with a delay, due to Government Digital Service needing to ramp up their work. Congruent with the central government's behaviour, local authorities also only started publishing agile related content after the local elections in 2018. Additionally, the spread of the methodology to the regional government might be explained by government experts from the field moving to the countryside in their mid-thirties/forties to have better "family life conditions", after gaining first work experience at GDS.

⁴⁰Link: <https://www.gov.uk/service-manual/communities/agile-delivery-community>

blogs of the respective organisations.⁴¹ Thus, it seems as if this medium has a special role to play in the information and discussion about these novel, agile working approaches – potentially, because a blog is a low key, less formal information channel where government officials dare to publish content related to less "traditional" topics.

In terms of average number of co-search terms, the Education & Skills Funding Agency and the Home Office lead the ranking with an equal average of 7 co-search terms per site, followed by the Department for Work & Pensions (5.4 terms/site; see Figure 11).⁴²

Figure 11: Average number of co-search terms per agil* site published by British government institutions

The final summarising remark for Britain's intra-country comparison is as follows: With Government Digital Service and the Cabinet Office there are clearly identifiable lead innovators with respect to the number of agil* sites published. Other institutions' efforts in that respect are substantially lesser. In particular, non-central government institutions (i.e. local authorities) started to publish agile content much later and with less intensity. This hints at a rather centralistic approach to introducing agile methods into government.

⁴¹The percentage was calculated by comparing blog sub-domains, e.g. "blog.gov.uk", with non-blog sub-domains, e.g. "gov.uk".

⁴²However, these results might be slightly skewed as the crawler only yielded few pages for the respective institutions (a total of 2, 3, and 5 pages respectively).

4.3 Inter-Country Comparison

In order to have the most legitimate basis for a quantitative comparison, only the central government ministerial departments are included in the following inter-country comparison.⁴³

Figure 12 depicts the wide lead – both in quantity and timing – that British institutions have over German ones in terms of agil* sites published: British central government institutions started publishing

Figure 12: Agil* sites published by British and German central government ministerial departments over time

agile content three years earlier and in total published 3.74 times more pages than German central government ministerial departments (85 British vs. 23 German sites).

As can be seen in Figure 13 (see next page), the most active British department in this respect – as well as by far the most active in general – is the Cabinet Office (37 sites), while in Germany it is the Federal Ministry of Labour and Social Affairs (13 sites).⁴⁴ In terms of highest number of average co-search terms per site, the German Federal Government Office (9 co-search terms on average) beats the British Home Office (7 terms). Yet, the crawler only yielded one page (German Federal Government) and three pages (British Home Office) for these institutions, and thus this high number might be skewed. Overall, it has to be noted that the total number of central government institutions publishing agile related content on their web presences is more than double for Britain (13 British vs. 6 German institutions), and the total average of agil* sites published (6.54 vs. 3.83 sites) as well as the total average number of co-search terms per site (3.7 vs. 2.36) is considerably higher for British institutions – see neon yellow markers in Figure 13).⁴⁵

⁴³The two countries' state systems are rather distinct – Britain is a unitary state, while Germany is federal (Elazar, 1997) – and hence, including non-central state institutions into the comparison was not reasonable. Furthermore, subordinate institutions were also excluded from the comparison due to their too country specific nature and number. Also, although the total number of British ministerial departments according to <https://www.gov.uk/government/organisations> is 24, this comparison only includes government bodies that are present in both countries – with the sole exception of the British Department for Digital, Culture Media & Sport. Finally, 22 of the British institutions' agil* sites – the majority of sites before 2015 – stem from gov.uk's national web archive. Germany does not have a comparable archive, and hence outdated and deleted German government pages mentioning agile methods might have simply been lost, potentially skewing the comparison in Britain's favour. Therefore, all British sites stemming from the web archive have been excluded in this analysis. As expected, including those pages further widens the lead of British institutions (see Figure 17 and 18 in Appendix A).

⁴⁴See Figure 6 in this section as well as Figure 16 in Appendix A for a depiction over time. Furthermore,

Figure 13: Breadth and depth of agil* sites published by British and German central government ministerial departments⁴⁶

When comparing the top frequent co-search terms appearing on all published agile related sites of British and German government institutions (see Figure 14 on next page), it becomes apparent that British institutions' content focuses way more on the user and their needs (three of the most frequent co-search terms relate to the user).⁴⁷ As opposed to this, on German central government agil* sites user ("Nutzer") only appears once amongst all co-search terms while most of the other more frequent terms are rather general and describe a new innovation culture ("innovationskultur") or a new agile way of working ("agile arbeitsformen" & "agil arbeit").⁴⁸ Furthermore, the frequent co-search terms suggest that Britain's descriptions are more hands on, as terms like "user research" or "service design" imply, while German institutions seem stuck on a meta level discussion about agile methods, without giving more detailed information about their particularities. The only exception are the co-search terms related to the scrum method (e.g. "scrum", "backlog", "sprint", and "product owner") which all stem from this one article of the Federal Gov-

see Figure 15 in Appendix A for a depiction of Figure 13 including the complete institution names.

⁴⁵As can be seen by the substantially lower respective numbers for other German institutions, the "high" total average number of co-search terms per site in the case of Germany is only driven by that one "outlier" Federal Government Office site – also, because the total number of government institutions publishing agile related content is lower and the respective numbers of the institutions have a stronger influence on the average. Hence, the difference between British and German government institutions in that respect is actually larger than the difference in averages suggests.

⁴⁶The German Ministry for Economic Affairs & Energy and the Ministry for the Environment, Nature Conservation & Nuclear safety yielded exactly the same results as can be seen by the blue-green triangle in the bottom left of the graph as well as in Figure 15 in the Appendix A.

⁴⁷For a depiction of the most frequent co-search terms as word clouds where the size of each word correlates with the number of times the co-search term appears on all agil* sites published by British and German (federal) government departments see Figure 19 in Appendix A.

⁴⁸For an English translation of all German co-search terms see Table 6 in Appendix B.

ernment Office already mentioned above. Generally, this is consistent with the rather "shallow" description of the topic by German institutions with regard to the average number of co-search terms (see Figure 13 above), as well as the comparatively lower frequent and more skewed distribution of co-search terms: Over 75% of all co-search terms appear only once on German central governments' agil* sites, and even the top co-search terms only appear on 13% of the sites (see Figure 14). It seems as if German central government institutions have not yet grasped the major difference when it comes to agile methods: The focus on users (i.e. citizens) and the fulfillment of their needs through thorough user research and good service design in iterative sprints (see Section 2.2). Much rather, they still seem to regard agile as a new, mostly non-public administration related trend in the world of work. The British co-search term frequencies, in contrast, suggest that the British institutions have done their homework when it comes to understanding the ins and outs of agile methods and presenting these on their web presences.

British Central Government Institutions

German Central Government Institutions

Figure 14: Frequency of co-search term lemmas for all agil* sites published by British and German central government ministerial departments

5 Discussion

5.1 Summary of Findings

This paper examined the evolution and spread of agile methods in British and German government institutions. To do so, the crawling results from government websites have been analysed for the appearance of agile terms as well as other topic relevant co-search terms.⁴⁹ Drawing on the results from Section 4, the major findings on a general level as well as with respect to an intra- and inter-country comparison are summarized in Table 4.

Table 4: Summary of findings

<i>Level</i>	Germany	United Kingdom
<i>General</i>	<ul style="list-style-type: none"> • Number of agile related sites published by government institutions as well as number of government domains publishing agile sites increased over time and particularly in recent years. • Most agil* sites published (84%) also mention digital transformation related keywords. • Both observations are in line with theses in agile related public management literature (cf. e.g. Mergel, Whitford et al., n.d.). 	
<i>Intra-Country (also including subordinate institutions as well as sites from gov.uk's web archive for Britain)</i>	<ul style="list-style-type: none"> • Agile is rather novel trend; first agil* sites published in 2015; 74 sites in total. • Most active publisher is Federal Ministry of Labour & Social Affairs (13 pp.) and the state of Baden-Wuerttemberg (9 pp.). • Federal Government Office (9 terms/site) and Bavarian government (3.2 terms/site) leading in terms of average number of co-search terms per site. • Average differences between central and non-central authorities with respect to timing and quantity of agil* sites published are minor. 	<ul style="list-style-type: none"> • Government institutions have been discussing agile methods on web presences since 2011; 380 published sites in total. • Leading publishers are Government Digital Service (143 pp.) and Cabinet Office (55 pp.). • Highest average number of co-search terms by Home Office and Education & Skills Funding Agency (7 terms/site each). • Average differences between central and non-central, i.e. local authorities are major.
<i>Inter-Country (only including central government ministerial departments)</i>	<ul style="list-style-type: none"> • British central government ministerial departments started publishing agil* sites 3 years earlier and published 3.7 times more sites than German ones (85 British vs. 23 German sites). • Number of British institutions publishing agile related content is more than twice as high (13 British vs. 6 German institutions); they published 1.71 times more sites on average (6.54 British vs. 3.83 German sites on average); and have a 1.57 times higher total average number of co-search terms per site (3.7 British vs. 2.36 German total average). 	

⁴⁹The average number of co-search terms gives an indication of how in-depth the topic is being discussed on the respective site.

*Inter-Country
(cont'd)*

- Most active British institution (Cabinet office, 37 sites) has 2.85 times more published agil* sites than most active German institution (Ministry of Labour & Social Affairs, 13 sites).
- German central government department with most in-depth description of method (Federal Government Office, 9 terms/site) has 1.29 times more co-search terms on average per site than respective British institution (Home Office, 7 terms/site) – yet, published only one site in total.
- While British central government ministerial departments show great knowledge about particularities of agile methods and care a lot about their users (high average frequency of co-search terms and most frequent terms are "user research" and "user need"), descriptions of agile methods on German institutions' web pages seem rather general, shallow, and discontinuous (lower co-search term frequency on average and most frequent terms are "innovation culture" and "agile ways of working").

The results can be read as sign for the increasing importance that agile methods hold for government institutions in the face of the digital transformation. Furthermore, they reveal the strong agile leadership position that British government institutions – in particular the ministerial organisations – have in comparison to Germany. Also, as mentioned in the Introduction, Government Digital Service – the centre of excellence for the British government's digital transformation – in fact turned out to have paved the way for a government-wide adoption of agile methods. Yet, the results also highlight the important role of the British Cabinet Office in that respect. Finally, the rather minor average differences between central and non-central government institutions in the case of Germany potentially hint at a somewhat federalistic approach to bringing agile methodologies into government. For Britain, the substantial differences between central and local authorities with respect to timing and quantity of agil* sites published suppose a rather centralistic approach.⁵⁰ Both these observations are in line with the unitary (Britain) or respectively federal (Germany) state system of each country (Elazar, 1997).

In addition to uncovering above findings, this work contributes to the literature by exemplifying how a "webometrical" analysis approach (Thelwall, 2009) based on self-crawled web content can yield substantive insights for public management research.

⁵⁰However, a direct quantitative comparison between Britain and Germany in that respect is not sensible since the German federal state institutions (still) own substantially more political leeway than British local authorities. (Despite attempts to improve the situation for British non-central bodies with the help of the Localism Act (legislation.gov.uk, 2011), the deep-rooted centralisation in Britain is hardly challenged (see for example Jeraj (2013) and Pipe (2013))).

5.2 Policy Implications

The policy implications of this work can be distinguished between implications addressing British government officials, implications for German government officials, and implications addressing officials of both and potentially also further countries.

For British Government officials the implications are rather complimentary. Their work, in particular the continuously persistent work of the Cabinet Office and Government Digital Service, appears to have paid off. According to the respective web publications, agile methods are spreading amongst a wide range of government institutions with their understanding of the methodology's particularities seeming rather substantial. Consequently, British government officials should keep up this spirit in order to truly and fully become an agile, digitally transformed government. One thing they could consider though, is strengthening local authorities in their application of agile methods. Since ultimately it is these government bodies that are responsible for providing a substantive set of vital services to the citizens (Local Government Association, 2020), and hence might also benefit from a more agile approach.

To German government officials, on the other hand, one would like to say: "You are late for the agile party and seem to have a rather unusual understanding of the party's theme." To catch up, they could possibly adopt some strategies of their British colleagues: First, a clear acknowledgement that the digital transformation of government needs agile methods to reach its full potential would set the necessary stage for public administrations to become active and stop seeing agile as a mostly external phenomenon. Perhaps they could do this by making the link between agile methods and digital transformation much clearer in their strategy papers on the topic (see Section 2.1). Second, they might consider having more distinct lighthouse institutions that internally push the application of agile methods and openly talk about it so that other government organisations can follow. Although the German Federal Ministry of Labour & Social Affairs seems to have opted for this position recently, a clearer sign from above, i.e. the Federal Government Office, might turn out to be more impactful.⁵¹ Third, they could establish more government related online blogs, which would provide a low key communication channel where government officials could share their experiences and knowledge on the topic.⁵² Fourth, in line with that,

⁵¹The latest efforts of the Federal Government Office (i.e. Tech and Work 4 Germany) are a good start but seem rather lukewarm and with limited radiant power. Whether or not this "sign from above" should even be a "lead from the top" as in Britain's case, and if so, whether this would be compatible with German's federal power dynamics is a question for another research paper.

⁵²So far, blogs are treated rather poorly by German government officials in that respect: Only two of the

establishing agile communities of practice can potentially provide a further useful platform for exchange. Fifth and finally, Germany could consider adopting a similar "outside-in-approach" as Britain: Their digital, and therewith agile transformation apparently strongly coincided with the implementation of the single government domain (gov.uk) by Government Digital Service (see Section 4.2.2). While this conversion potentially provided a much more efficient and effective way to reach out to citizens and fulfill their needs, it also made government institutions come into contact with Government Digital Service when transferring their web content to the new domain, and thereby might also have provided them with a first possibility to experience the advantages of using agile methods.⁵³

Finally, the policy implication for British as well as German government officials, and potentially also beyond, is that agile methods apparently have become a considerable tool for managing projects related to government's digital transformation. Hence, further experimenting and honest best practice sharing will be needed in order to fully reap the methodology's potential in better serving the citizens' needs with digital services.

5.3 Limitations & Future Work

This work is not without limitations; they come with suggestions for future research. First, the web crawl only included plain HTML web site content, and hence information from PDF, video or audio files for example was not analysed. Since, for certain government institutions uploading information in PDF format (still) might be common practice, future iterations should potentially try to find ways to also include these type of online sources. Second, due to time limitations, the whole process of data collection as well as all the design choices for data preprocessing were not analysed as thoroughly as they potentially could have been.⁵⁴ Future work thus might dive deeper into the log files of such crawls to identify potential subtle irregularities that might impede the reliability of the crawled data or compare various data cleaning approaches in terms of their accuracy. Third, following this line of thought, future researchers could build their own crawler architecture from scratch to not rely on third party code (in this case the Scrapy package), and further tailor the crawler's behaviour to their research goal. Fourth, the composition and analysis

crawler yielded agil* sites, as opposed to more than half of the British sites, originated from government blogs (Link to blog entry one and link to blog entry two, in addition both entries are only rather loosely related to government institutions).

⁵³Interestingly, the most active German federal states in terms of number of agil* sites published and average number of co-search terms (Baden-Wuerttemberg, Bavaria, and Hesse) all have (mostly) single government domains (only two ministries from Baden-Wuerttemberg have their own domains).

⁵⁴Besides, of course, all the analyses to avoid obviously identifiable errors.

of the co-search terms could potentially take a more data driven approach based on the gathered texts themselves, e.g. by analysing all words in a certain window around the *agil** term, and not only rely upon manually predefined dictionaries. Fifth, the seed URL for Britain was set to *gov.uk*, which basically includes the web presences of all British government institutions and agencies. However, for Germany, which does not have a single government domain, only the ministerial websites have been set as seed URLs, thus leading to an exclusion of potential subordinate authorities like the IT provider for the federal government, ITZBund. Hence, in order to also yield insights on the role of such subordinate institutions and further expand the basis for a legitimate inter-country comparison, future research could include their respective web presences as additional seed URLs. Finally, since the crawler yielded pages can only serve as a proxy for the true status of the application of agile methods in the respective institutions, researchers could potentially triangulate these results with supplementary data sources, for example qualitative interviews, quantitative surveys, further online data etc.

Additional research alleys that could be build on or amend the results of this paper, but go beyond the core of its research question, are as follows. Researchers could dive deeper into analysing the text content with author or plagiarism detection methods in order to find out which author(s) or institution(s) actually "feed" the appearance of agile methods on government websites. Also, web sites from NGOs, consultancies or relevant media outlets could be included in the crawls to inspect the relation between these sites and government websites, and potentially find out who or what "drives" the adoption of agile methods in the public sector. Furthermore, future work could examine how the differences in state system – federal Germany vs. unitary Britain – influence the adoption of agile methods or digital transformation in general. Finally, researchers could further explore how crawled online content can serve as reliable data source for social science research.

6 Conclusion

By analysing the appearance of agile related keywords on institutional web sites over time, this study explored the evolution and spread of agile methods in the British and German government. While the two countries' institutions share some common features like an overall increasing publication trend of agile sites and a majority of these sites also mentioning terms connected to digitalisation, the numbers clearly identify the Britons as the leading agile innovator: Not only have they started publishing agile related content many years before their German peers, the spread of the methodology amongst the studied government institution is also wider, and the respective method descriptions are considerably more detailed. Although it turned out that the British government's digital service had played the leading role, the Cabinet Office's promotion of agile methods also proved to be very substantial. In order to catch up, German government could apply some of the leader's strategies like giving the methodology more weight in strategic papers on digital transformation or establishing agile lighthouse institutions.

References

- A. Salloum, S., et al. (2017). Mining Social Media Text: Extracting Knowledge from Facebook. *International Journal of Computing and Digital Systems*. <https://doi.org/10.12785/ijcds/060203>
- Agile Manifesto. (2001). Manifesto for Agile Software Development. <http://agilemanifesto.org/>
- Altukhova, N., Vasileva, E. & Slavin, B. (2016). Concept for a new approach to project management in the activities of public servants. *Business Informatics*. <https://doi.org/10.17323/1998-0663.2016.4.60.69>
- Andriole, S. J. (2018). Skills and Competencies for Digital Transformation. *IT Professional*, 20(6). <https://doi.org/10.1109/MITP.2018.2876926>
- Australian Digital Transformation Agency. (2019). Agile and User-Centred Processes. <https://www.dta.gov.au/help-and-advice/digital-service-standard/digital-service-standard-criteria/3-agile-and-user-centred-process>
- Barbarese, A. (2015). *Ad hoc and general-purpose corpus construction from web sources* (Doctoral dissertation). ENS Lyon. <https://doi.org/10.1167309>
- Barbarese, A. (2020). Adbar/htmldate: Find original and updated publication dates of web pages using common patterns, heuristics and robust extraction. <https://zenodo.org/record/3611241#.XlgETRP0mhc>
- Behörden Spiegel. (2020). Positives Menschenbild – Basis für digitale Transformation. <https://www.behörden-spiegel.de/2020/01/17/positives-menschenbild-basis-fuer-digitale-transformation/>
- Berger, H. (2007). Agile development in a bureaucratic arena-A case study experience. *International Journal of Information Management*. <https://doi.org/10.1016/j.ijinfomgt.2007.08.009>
- Berger, H. Et al. (2005). A UK Government IS Project: Inherent Cultural Issues of a Bureaucratic Environment Impacting upon a RAD-type Development, In *Pacific asia conference on information systems 2005, sections 1–8 and poster sessions 1–6, kaohsiung*.
- Beyer, H. (2010). User-Centered Agile Methods. *Synthesis Lectures on Human-Centered Informatics*. <https://doi.org/10.2200/s00286ed1v01y201002hci010>
- Bloomberg, J. (2018). Digitization, Digitalization, And Digital Transformation: Confuse Them At Your Peril. *Forbes Media LLC*.
- Boehmer, A., Hostettler, R., Richter, C., Lindemann, U., Conradt, J. & Knoll, A. (2017). Towards Agile Product Development - The Role of Prototyping. *Proceedings of the International Conference on Engineering Design, ICED*.
- Branco, M. C. & Rodrigues, L. L. (2006). Communication of corporate social responsibility by Portuguese banks: A legitimacy theory perspective. *Corporate Communications*. <https://doi.org/10.1108/13563280610680821>
- Brennen, S. & Kreiss, D. (2016). Digitalization (K. Jensen, E. Rothenbuhler, J. Pooley & R. Craig, Eds.). In K. Jensen, E. Rothenbuhler, J. Pooley & R. Craig (Eds.), *The international encyclopedia of communication theory and philosophy*. Wiley.
- Bundesministerium des Inneren für Bau und Heimat. (2017). Onlinezugangsgesetz (OZG). <https://www.bmi.bund.de/DE/themen/moderne-verwaltung/verwaltungsmodernisierung/onlinezugangsgesetz/onlinezugangsgesetz-artikel.html>
- Cabinet Office of the United Kingdom. (2017). Government Transformation Strategy. <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy>

- Carvalho Fernandes, M., Juarez Alencar, A., Assis Schmitz, E., Ferreira da Silva, M. & Sotirios Stefaneas, P. (2016). Evaluation of Agile Software Projects in the Public Sector: A Literature Review. *Journal of Software*. <https://doi.org/10.17706/jsw.11.3.312-325>
- Clark, J. & DeRose, S. (1999). XML Path Language (XPath). *W3C Recommendation*.
- Clarke, A. (2019). Digital Government Units: What Are They, and What Do They Mean for Digital Era Public Management Renewal? *International Public Management Journal*. <https://doi.org/10.1080/10967494.2019.1686447>
- Cothey, V. (2004). Web-Crawling Reliability. *Journal of the American Society for Information Science and Technology*. <https://doi.org/10.1002/asi.20078>
- Cunnington, K. (2017). The future of digital government. <https://civilservice.blog.gov.uk/2017/02/09/the-future-of-digital-government/>
- Da Silva, T. S., Martin, A., Maurer, F. & Silveira, M. (2011). User-centered design and agile methods: A systematic review, In *Proceedings - 2011 agile conference, agile 2011*. <https://doi.org/10.1109/AGILE.2011.24>
- Deutscher Bundestag. (2017). Gesetz zur Verbesserung des Onlinezugangs zu Verwaltungsleistungen (Onlinezugangsgesetz - OZG). <https://www.gesetze-im-internet.de/ozg/OZG.pdf>
- Dierks, M. (2020). Mut und neues Mindset - Digitale Transformation und Partizipation. *Behörden Spiegel*, 28.
- Dingsøyr, T., Nerur, S., Balijepally, V. & Moe, N. B. (2012). A decade of agile methodologies: Towards explaining agile software development. *Journal of Systems and Software*, 85(6), 1213–1221. <https://doi.org/10.1016/j.jss.2012.02.033>
- Dubinsky, Y., Talby, D., Hazzan, O. & Keren, A. (2005). Agile metrics at the Israeli Air Force, In *Proceedings - agile confernce 2005*. <https://doi.org/10.1109/ADC.2005.8>
- Eby, K. (2016). Comprehensive Guide to the Agile Manifesto. <https://www.smartsheet.com/comprehensive-guide-values-principles-agile-manifesto>
- Elazar, D. J. (1997). Contrasting unitary and federal systems. *International Political Science Review*. <https://doi.org/10.1177/019251297018003002>
- European Commission. (2017a). Ministerial Declaration on eGovernment - the Tallinn Declaration. <https://ec.europa.eu/digital-single-market/en/news/ministerial-declaration-egovernment-tallinn-declaration>
- European Commission. (2017b). *Tallinn Declaration on eGovernment* (tech. rep.). http://ec.europa.eu/newsroom/document.cfm?doc_id=47559
- European Commission. (2019a). *Digital Government Factsheet 2019 - Germany* (tech. rep.). https://joinup.ec.europa.eu/sites/default/files/inline-files/Digital_Government_Factsheets_Germany_2019.pdf
- European Commission. (2019b). *Digital Government Factsheet 2019 - The United Kingdom* (tech. rep.). https://joinup.ec.europa.eu/sites/default/files/inline-files/Digital_Government_Factsheets_UK_2019.pdf
- European Union. (2018). REGULATION (EU) 2018/1724 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 2 October 2018 establishing a single digital gateway to provide access to information, to procedures and to assistance and problem-solving services and amending Regulation (EU). Official Journal of the European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1724&from=EN>
- Faassen, M. (2006). lxml - XML and HTML with Python. <https://lxml.de/>
- Fridman, A. (2016). The Massive Downside of Agile Software Development. *Inc.*
- Fruhling, A., McDonald, P. & Dunbar, C. (2008). A case study: Introducing eXtreme programming in a US government system development project, In *Proceedings of*

- the annual hawaii international conference on system sciences*. <https://doi.org/10.1109/HICSS.2008.4>
- Fulgham, C., Johnson, J., Crandall, M., Jackson, L. & Burrows, N. (2011). The FBI gets agile. *IT Professional*. <https://doi.org/10.1109/MITP.2011.88>
- GeoPandas Developers. (2013). GeoPandas 0.7.0. <https://geopandas.org/>
- Ghosh, T., Anderson, S. J., Elvidge, C. D. & Sutton, P. C. (2013). Using nighttime satellite imagery as a proxy measure of human well-being. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su5124988>
- Gillies, S. (2020). descartes 1.1.0. <https://pypi.org/project/descartes/>
- Goldman, S. L., Nagel, R. N. & Preiss, K. (1994). *Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer*. San Francisco, CA, Wiley.
- Government Digital Service. (2020). About the Government Digital Service. <https://gds.blog.gov.uk/about/>
- Greve, C., Ejersbo, N., Lægreid, P. & Rykkja, L. H. (2019). Unpacking Nordic Administrative Reforms: Agile and Adaptive Governments. *International Journal of Public Administration*. <https://doi.org/10.1080/01900692.2019.1645688>
- Hajjdiab, H. & Taleb, A. S. (2011). Adopting Agile Software Development: Issues and Challenges. *International Journal of Managing Value and Supply Chains*. <https://doi.org/10.5121/ijmvsc.2011.2301>
- Henriette, E., Feki, M. & Boughzala, I. (2015). The Shape of Digital Transformation: A Systematic Literature Review. *Mediterranean Conference on Information Systems (MCIS) Proceedings*.
- Honnibal, M. & Johnson, M. (2015). An Improved Non-monotonic Transition System for Dependency Parsing, In *Proceedings of the 2015 conference on empirical methods in natural language processing*, Lisbon, Portugal, Association for Computational Linguistics. <https://aclweb.org/anthology/D/D15/D15-1162>
- Hughes, R. (2013). Iterative Development in a Nutshell, In *Agile data warehousing project management*. <https://doi.org/10.1016/b978-0-12-396463-2.00002-8>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*. <https://doi.org/10.1109/MCSE.2007.55>
- Ignatow, G. & Mihalcea, R. (2018). *Text Mining: A Guidebook for the Social Sciences*. <https://doi.org/10.4135/9781483399782>
- Iliev, M., Krasteva, I. & Ilieva, S. (2009). A case study on the adoption of measurable agile software development process.
- Jaeger, R. G. & Halliday, T. R. (1998). On Confirmatory versus Exploratory Research. *Herpetologica*.
- Jeraj, S. (2013). The jury is still out on the general power of competence.
- Kannan, V., Smita, J. & Verma, S. (2014). Agile vs waterfall: A Comparative Analysis. *International Journal of Science, Engineering and Technology Research (IJSETR)*, 3(10), 2680–2686.
- Khalil, S. & Fakir, M. (2017). RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*. <https://doi.org/10.1016/j.softx.2017.04.004>
- Kouzis-Loukas, D. (2016). Learning Scrapy. Packt Publishing Ltd.
- Kusay-Merkle, U. (2018). *Agiles Projektmanagement im Berufsalltag*. <https://doi.org/10.1007/978-3-662-56800-2>
- legislation.gov.uk. (2011). Localism Act 2011. <http://www.legislation.gov.uk/ukpga/2011/20/contents/enacted/data.htm>
- Liu, B. & Zhang, L. (2012). A survey of opinion mining and sentiment analysis, In *Mining text data*. <https://doi.org/https://doi.org/10.1007/978-1-4614-3223-4>

- Local Government Association. (2020). What is local government? <https://www.local.gov.uk/about/what-local-government>
- Lüdeling, A., Evert, S. & Baroni, M. (2015). Using web data for linguistic purposes (M. Hundt, N. Nadja & C. Biewe, Eds.). In M. Hundt, N. Nadja & C. Biewe (Eds.), *Corpus linguistics and the web*. <https://doi.org/10.1163/9789401203791>
- Lundqvist, M. & Olesen, P. B. (2016). *From waterfall to agile: How a public-sector agency successfully changed its system- development approach to become digital* (tech. rep.). McKinsey & Company. <http://www.mckinsey.com/business-functions/business-technology/our-insights/Digitizing-the-delivery-of-government-services?cid=other-eml-alt-mip-mck-oth-1603>
- Macaya, J. F. M., Alves, A. F., Meirelles, F. & Cunha, M. A. (2019). Digital by Default: The Use of Service Channels by Citizens, In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. https://doi.org/10.1007/978-3-030-27325-5_{_}25
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*.
- McKinsey. (2018). Unlocking success in digital transformations. <https://www.mckinsey.com/business-functions/organization/our-insights/unlocking-success-in-digital-transformations>
- McMahon, P. E. (2006). Lessons learned using agile methods on large defense contracts. *CrossTalk*.
- Mergel, I. (2016). Agile innovation management in government: A research agenda. *Government Information Quarterly*, 33(3), 516–523. <https://doi.org/10.1016/j.giq.2016.07.004>
- Mergel, I. (2019). Digitale Transformation als Reformvorhaben der deutschen öffentlichen Verwaltung. *der moderne staat – Zeitschrift für Public Policy, Recht und Management*. <https://doi.org/10.3224/dms.v12i1.09>
- Mergel, I., Edelmann, N. & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly*, 36(4). <https://doi.org/10.1016/j.giq.2019.06.002>
- Mergel, I., Gong, Y. & Bertot, J. (2018). Agile government: Systematic literature review and future research. *Government Information Quarterly*, 35(2), 291–298. <https://doi.org/10.1016/j.giq.2018.04.003>
- Mergel, I., Whitford, A. B. & Ganapati, S. (n.d.). Agile: A new way of Governing. *Public Administration Review*, to appear. <https://www.polver.uni-konstanz.de/en/mergel/open-public-administration-scholarship/publications/publications-in-progress/usercentereddesign-friedrichshafen/>
- Moore, R. J. (2001). Evolving to a "lighter" software process: A case study, In *26th annual nasa goddard software engineering workshop, ieee/nasa sew 2001*. <https://doi.org/10.1109/SEW.2001.992650>
- Mueller, A. (2020). wordcloud. http://amueller.github.io/word_cloud/
- Neuwirth, E. (2020). RColorBrewer. <https://www.rdocumentation.org/packages/RColorBrewer/versions/1.1-2/topics/RColorBrewer>
- Niemeyer, G. (2003). Dateutil - powerful extensions to datetime. <https://github.com/dateutil/dateutil>
- Nuottila, J., Aaltonen, K. & Kujala, J. (2016). Challenges of adopting agile methods in a public organization. *International Journal of Information Systems and Project Management*, 4(3), 65–85. <https://doi.org/10.12821/ijispm040304>
- Olston, C. & Najork, M. (2010). Web crawling. *Foundations and Trends in Information Retrieval*. <https://doi.org/10.1561/15000000017>

- Osmundsen, K. (2020). Competences for Digital Transformation: Insights from the Norwegian Energy Sector, In *Proceedings of the 53rd hawaii international conference on system sciences*.
- Parcell, J. & Holden, S. H. (2013). Agile policy development for digital government: An exploratory case study. *ACM International Conference Proceeding Series*, (April), 11–17. <https://doi.org/10.1145/2479724.2479731>
- Peterjohann, H. (2020). Scrum-Glossar. <https://www.scrum-glossar.de/>
- Pilgrim, M., Blanchard, D. & Cordasco, I. (2015). chardet. <https://chardet.readthedocs.io/en/latest/index.html>
- Pipe, J. (2013). Two years on, what has the Localism Act achieved? <https://www.theguardian.com/local-government-network/2013/nov/02/localism-act-devolution-uk-local-authorities>
- Richardson, L. (2007). Beautiful Soup Documentation. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- Rigby, D. K., Sutherland, J. & Takeuchi, H. (2016a). Embracing agile. *Harvard Business Review*.
- Rigby, D. K., Sutherland, J. & Takeuchi, H. (2016b). The Secret History of Agile Innovation. *Harvard Business Review*.
- Schäfer, R. & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, (2006), 486–493.
- Schwaber, K. & Sutherland, J. (2020). Scrum Guides.
- Scrapy. (2018). Scrapy 1.8 Documentation. <https://docs.scrapy.org/en/latest/index.html>
- Scrum.org. (2019). What is Scrum? A Better Way Of Building Products. <https://www.scrum.org/resources/what-is-scrum>
- Sherrell, L. (2013). Waterfall Model, In *Encyclopedia of sciences and religions*. <https://doi.org/10.1007/978-1-4020-8265-8>
- Shumate, M. & Weber, M. S. (2016). The Art of Web Crawling for Social Science Research (E. Hargittai & C. Sandvig, Eds.). In E. Hargittai & C. Sandvig (Eds.), *Digital research confidential: The secrets of studying behavior online*. MIT Press.
- Sivarajah, U., Irani, Z. & Jones, S. (2014). Application of web 2.0 technologies in e-government: A United Kingdom case study, In *Proceedings of the annual hawaii international conference on system sciences*. <https://doi.org/10.1109/HICSS.2014.280>
- Sørensen, E. & Torfing, J. (2011). Enhancing collaborative innovation in the public sector. *Administration and Society*. <https://doi.org/10.1177/0095399711418768>
- Sun, Y., Zhuang, Z. & Giles, C. L. (2007). A large-scale study of robots.txt, In *16th international world wide web conference, www2007*. <https://doi.org/10.1145/1242572.1242726>
- Surdu, J. & Parsons, D. J. (2006). Army simulation program balances agile and traditional methods with success.
- Tabrizi, B., Lam, E., Girard, K. & Irvin, V. (2019). Digital transformation is not about technology. *Harvard Business Review*.
- Takeuchi, H. & Nonaka, I. (1986). The New New Product Development Game. *Harvard Business Review*.
- Tang, H.-j., Yan, D.-f. & Tian, Y. (2013). Semantic dictionary based method for short text classification. *The Journal of China Universities of Posts and Telecommunications*, 20(1), 15–19. [https://doi.org/10.1016/S1005-8885\(13\)60256-3](https://doi.org/10.1016/S1005-8885(13)60256-3)

- Thelwall, M. (2009). Introduction to Webometrics: Quantitative Web Research for the Social Sciences. *Synthesis Lectures on Information Concepts, Retrieval, and Services*. <https://doi.org/10.2200/s00176ed1v01y200903icr004>
- Tridgell, A. & Mackerras, P. (2020). rsync(1) - Linux man page. <https://linux.die.net/man/1/rsync>
- Upender, B. (2005). Staying agile in government software projects, In *Proceedings - agile confrence 2005*. <https://doi.org/10.1109/ADC.2005.41>
- Vacari, I. & Prikladnicki, R. (2015). Adopting agile methods in the public sector: A systematic literature review. *Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE, 2015-Janua*, 709–714.
- Van Rossum, G. & Drake Jr, F. L. (1995). *Python Tutorial*. Amsterdam, The Netherlands, Centrum voor Wiskunde en Informatica.
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *Journal of Strategic Information Systems*, 28(2), 118–144. <https://doi.org/10.1016/j.jsis.2019.01.003>
- w3schools.com. (2020). XPath Syntax. https://www.w3schools.com/xml/xpath_syntax.asp
- Wang, C., Zhao, S., Kalra, A., Borcea, C. & Chen, Y. (2019). Webpage Depth Viewability Prediction Using Deep Sequential Neural Networks. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/TKDE.2018.2839599>
- Waskom, M. (2012). seaborn: statistical data visualization. <https://seaborn.pydata.org/>
- Wirdemann, R., Mainusch, J., Wirdemann, R. & Mainusch, J. (2017). User Stories, In *Scrum mit user stories*. <https://doi.org/10.3139/9783446450776.004>
- Xie, K., Di Tosto, G., Lu, L. & Cho, Y. S. (2018). Detecting leadership in peer-moderated online collaborative learning through text mining and social network analysis. *Internet and Higher Education*. <https://doi.org/10.1016/j.iheduc.2018.04.002>

Appendix A

Figure 15: Breadth and depth of agil* sites published by British and German central government ministerial departments (with institution names)

Figure 16: Agil* sites published by British central government ministerial departments over time (excluding sites from gov.uk's national web archive)

Figure 17: Agil* sites published by British and German central government ministerial departments over time (including sites from gov.uk's national web archive)

Figure 18: Breadth and depth of agil* sites published by British and German central government ministerial departments (with institution names and including sites from gov.uk's national web archive)

Figure 19: Frequency of co-search term lemmas for all agil* sites published by British and German central government ministerial departments depicted as word clouds

Appendix B

Table 5: List of government websites

Country / Level / Ministry	Source / Website
United Kingdom – Federal & State Level	https://www.gov.uk/government/organisations
All institutions are reachable under one domain	https://www.gov.uk
Germany – Federal Level	https://www.protokoll-inland.de/Webs/PI/DE/rang-titulierung/amtliche-reihenfolgen/bundesministerien/liste-der-bundesministerien-node.html
Auswärtiges Amt	https://www.auswaertiges-amt.de
Bundesministerium der Finanzen	https://www.bundesfinanzministerium.de
Bundesministerium der Justiz und für Verbraucherschutz	https://www.bmjv.de
Bundesministerium der Verteidigung	https://www.bmvg.de
Bundesministerium des Innern, für Bau und Heimat	https://www.bmi.bund.de
Bundesministerium für Arbeit und Soziales	https://www.bmas.de
Bundesministerium für Bildung und Forschung	https://www.bmbf.de
Bundesministerium für Ernährung und Landwirtschaft	https://www.bmel.de
Bundesministerium für Familie, Senioren, Frauen und Jugend	https://www.bmfsfj.de
Bundesministerium für Gesundheit	https://www.bundesgesundheitsministerium.de
Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit	https://www.bmu.de
Bundesministerium für Verkehr und digitale Infrastruktur	https://www.bmvi.de
Bundesministerium für Wirtschaft und Energie	https://www.bmwi.de
Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung	https://www.bmz.de
Bundesregierung	https://www.bundesregierung.de
Germany – State Level	
<i>Baden-Wuerttemberg</i>	https://www.baden-wuerttemberg.de/de/regierung/landesregierung/ministerien
Ministerium der Justiz und für Europa	https://www.justizministerium-bw.de
Ministerium für Kultus, Jugend und Sport	https://www.km-bw.de
All other institutions are reachable under one domain	https://www.baden-wuerttemberg.de
<i>Bavaria</i>	https://www.bayern.de/staatsregierung/staatsministerien
All institutions are reachable under one domain	https://www.bayern.de
<i>Berlin</i>	https://www.berlin.de/rbmskzl/regierender-buergermeister/senat

Senatsverwaltung Stadtentwicklung und Wohnen	für	https://www.stadtentwicklung.berlin.de
All other institutions are reachable under one domain		https://www.berlin.de
<i>Brandenburg</i>		https://service.brandenburg.de/lis/list.php?page=behoerdenverzeichnis_art&sv[adr_art]=lb_Min&sort=org_name1,org_name2,org_name3,org_name4&_grid=Ministerien
All institutions are reachable under one domain		https://www.brandenburg.de
<i>Bremen</i>		https://landesportal.bremen.de/senat/ressorts
All institutions are reachable under one domain		https://www.bremen.de
<i>Hamburg</i>		https://www.hamburg.de/behoerden
All institutions are reachable under one domain		https://www.hamburg.de
<i>Hesse</i>		https://www.hessen.de/regierung/staatskanzlei-und-ministerien
All institutions are reachable under one domain		https://www.hessen.de
<i>Lower Saxony</i>		https://www.niedersachsen.de/politik_staatslandesregierung_ministerien/die-niedersaechsische-landesregierung-20076.html
All institutions are reachable under one domain		https://www.niedersachsen.de
<i>Mecklenburg-Vorpommern</i>		https://www.regierung-mv.de/
All institutions are reachable under one domain		https://www.regierung-mv.de
<i>North Rhine-Westphalia</i>		https://www.land.nrw/de/ministerien-und-vertretungen
Ministerium der Finanzen		http://www.fm.nrw.de
Ministerium der Justiz		https://www.justiz.nrw
Ministerium des Innern		http://www.im.nrw.de
Ministerium für Arbeit, Gesundheit und Soziales		https://www.mags.nrw
Ministerium für Bundes- und Europaangelegenheiten sowie Internationales		https://www.mbei.nrw
Ministerium für Heimat, Kommunales, Bau und Gleichstellung		https://www.mhkgb.nrw
Ministerium für Kinder, Familie, Flüchtlinge und Integration		https://www.mkffi.nrw
Ministerium für Kultur und Wissenschaft		https://www.mkw.nrw
Ministerium für Schule und Bildung		https://www.schulministerium.nrw.de
Ministerium für Umwelt, Landwirtschaft, Natur- und Verbraucherschutz		https://www.umwelt.nrw.de
Ministerium für Verkehr		http://www.vm.nrw.de
Ministerium für Wirtschaft, Innovation, Digitalisierung und Energie		https://www.wirtschaft.nrw
Staatskanzlei		https://www.land.nrw
<i>Rhineland-Palatinate</i>		https://www.rlp.de/de/landesregierung/ministerien

All institutions are reachable under one domain	https://www.rlp.de
<i>Saarland</i>	https://www.saarland.de/12341.htm
All institutions are reachable under one domain	https://www.saarland.de
<i>Saxony</i>	https://www.sachsen.de/regierung-verwaltung-egovernment.html
All institutions are reachable under one domain	https://www.sachsen.de
<i>Saxony-Anhalt</i>	https://www.sachsen-anhalt.de/lj/politik-und-verwaltung/die-landesregierung/ministerien/ministerien
All institutions are reachable under one domain	https://www.sachsen-anhalt.de
<i>Schleswig-Holstein</i>	https://www.schleswig-holstein.de/DE/Landesregierung/landesregierung_node.html
All institutions are reachable under one domain	https://www.schleswig-holstein.de
<i>Thuringia</i>	https://www.thueringen.de
Thüringer Ministerium für Arbeit, Soziales, Gesundheit, Frauen und Familie	https://www.tmasgff.de
Thüringer Staatskanzlei	https://www.staatskanzlei-thueringen.de
All other institutions are reachable under one domain	https://www.thueringen.de

Table 6: List of English and German co-search terms

German	English
abteilungsübergreifender Ansatz	cross-departmental approach
agil arbeitende Organisationseinheit	agile organization unit
agil arbeitende Teams	agile team (<i>double</i>)
agile Arbeitsformen	agile working methods (<i>double</i>)
agile Arbeitskultur	agile working culture
agile Arbeitsmethoden	agile working methods
agile Arbeitspraktiken	agile work practices
agile Arbeitsstruktur	agile working structure
agile Arbeitsumgebung	agile environment
agile Arbeitsweise	agile way of working
agile Arbeitswelt	agile working environment (<i>deleted because of confusion with flexible working arrangements - agile working</i>)
agile Arbeits- und Managementmethoden	agile working and management methods
agile Arbeits- und Teamkultur	agile work and team culture
agile Bereitstellung	agile delivery
agile digitale Fähigkeit	agile digital capability
agile Entwicklung	agile development
agile Fähigkeiten	agile capability
agile Herangehensweise	agile approach (<i>deleted because of too many appearances in non-related contexts</i>)
agile Innovation	agile innovation
agile Kultur	agile culture
agile Managementmethoden	agile management methods
agile Methode	agile method
agile Methodik	agile methodology
agile Organisation	agile organisation
agile Phasen	agile stages
agile Prinzipien	agile principles
agile Prozesse	agile processes
agile Softwareentwicklung	agile software development
agile Strukturen	agile structures
agile Techniken	agile techniques
agile Transformation	agile transformation
agile Vorgehensweise	agile procedures
agile Wände	agile walls
agile Weise	agile way (<i>deleted because of too many appearances in non-related contexts</i>)
agile Zeremonie	agile ceremony

agile Zertifizierung	agile certification
agiler Ansatz	agile approach (<i>deleted because of too many appearances in non-related contexts</i>)
agiler Coach	agile coach
agiler Prozess	agile process
agiler Wandel	agile transition
agiler Wert	agile Value
agiles Arbeiten	agile work (<i>deleted because of confusion with flexible working arrangements - agile working</i>)
agiles Betriebsmodell	agile operating model
agiles Coaching	agile coaching
agiles Management	agile management
agiles Manifest	agile manifesto
agiles Mindset	agile mindset
agiles Programmieren	agile programming
agiles Projekt	agile project
agiles Projektmanagement	agile project management
agiles Team	agile team
agiles Training	agile training (<i>deleted because of too many appearances in non-related contexts</i>)
agiles Zertifikat	agile certificate
Backlog	backlog
Daily Scrum	daily scrum
Definition of Done	definition of done
Definition of Ready	definition of ready
Design Thinking	design thinking
Empathy Map	empathy map
Entwicklungsteam	development team
Extremnutzer	extreme user
fail fast	fail fast
fail forward	fail forward
frühes Scheitern	fail early
Hackathon	hackathon
hierarchiefrei	free of hierarchy
hierarchiefreier Raum	hierarchy free space
Innovationskultur	innovation culture
Innovationsmanagement	innovation management
Iteration	iteration
iterativ	iterative
Kanban	kanban

Kanban Board	kanban board
kollaborativ	collaborative (<i>deleted because of too many appearances in wrong context</i>)
Kollaborationsräume	collaboration spaces
kollaboratives Design	collaborative design
kurzzyklische Ergebnisse	short-cycle results
Lean Start up	lean start up
Lean Start-up	lean start-up
Lean Startup	lean startup
Minimum Viable Product	minimum viable product
MVP	MVP
neue Arbeitsmodelle	new working models
neue Arbeitsweisen	new working methods
neue Team- und Führungsstrukturen	new team and management structures
New Work	new work (<i>deleted since term seems to be only used in Germany</i>)
Nutzer	user (<i>deleted because of too many appearances in wrong context</i>)
Nutzeranforderungen	user requirements
Nutzerbedürfnisse	user needs
Nutzerforscher	user researcher
Nutzerforschung	user research
nutzerorientiert	user-oriented
nutzerorientierte Lösungen	user-friendly solutions
Nutzerstory	user story
Nutzertest	user testing
nutzerzentriert	user-centred
nutzerzentriertes Design	user-centred design
Planning Poker	planning poker
Product Backlog	product backlog
Product Owner	product owner
Project Canvas	project canvas
schlanke Softwareentwicklung	lean software development
Scrum	scrum
Scrum Coach	scrum coach
Scrum Master	scrum master
Scrum Team	scrum team
Service Design	service design
Softwareentwicklung	software development
Sprint	sprint
Sprint Backlog	sprint backlog

Sprint Planning	sprint planning
Sprint Retrospective	sprint retrospective
Sprint Review	sprint review
Timebox	time box
Timeboxing	timeboxing
visuelles Denken	visual thinking
Whiteboard	whiteboard
Wicked Problem	wicked problem
Wireframe	wireframe

Appendix C

The general work flow of the Scrapy architecture is as follows (see Figure 20; adapted from "Architecture overview" in Scrapy, 2018):

1. The Engine gets the initial Requests to crawl from the Spider.
2. The Engine schedules the Requests in the Scheduler and asks for the next Requests to crawl.
3. The Scheduler returns the next Requests to the Engine.
4. The Engine sends the Requests to the Downloader, passing through the Downloader Middlewares.
5. Once the page finishes downloading the Downloader generates a Response (with that page) and sends it to the Engine, passing through the Downloader Middlewares.
6. The Engine receives the Response from the Downloader and sends it to the Spider for processing, passing through the Spider Middleware.
7. The Spider processes the Response and returns scraped items and new Requests (to follow) to the Engine, passing through the Spider Middleware.
8. The Engine sends processed items to Item Pipelines, then sends processed Requests to the Scheduler and asks for possible next Requests to crawl.
9. The process repeats (from step 1) until there are no more requests from the Scheduler.

Figure 20: Scrapy Architecture (Source: <https://docs.scrapy.org>, reprinted with permission)

Table 7: Number of sites per state / country before and after preprocessing

	Number of Crawler Retrieved Sites	Number of Sites after Preprocessing / Cleaning
United Kingdom	6185	380
Germany (total)	165384	71
Germany Federal Level	200	23
Baden-Wuerttemberg	130	9
Bavaria	230	5
Berlin	70543	8
Brandenburg	169	1
Bremen	1329	6
Hamburg	37	2
Hesse	64	4
Lower Saxony	0	0
Mecklenburg-Vorpommern	4	1
North Rhine-Westphalia	113	0
Rhineland-Palatinate	167	6
Saarland	59290	0
Saxony	33044	0
Saxony-Anhalt	21	3
Schleswig-Holstein	31	1
Thuringia	12	2

Statement of Authorship

I hereby confirm and certify that this master thesis is my own work. All ideas and language of others are acknowledged in the text. All references and verbatim extracts are properly quoted and all other sources of information are specifically and clearly designated.

DATE: 25th of May 2020

NAME: Maximilian Kupi

SIGNATURE:

A handwritten signature in black ink, appearing to read 'MKupi', with a stylized, flowing script.