Manchester Metropolitan University

e-space

Manchester Metropolitan University's Research Repository

Please cite the Published Version

Rainer, Austen and Williams, Ashley (2019) Using blog-like documents to investigate software practice: Benefits, challenges, and research directions. Journal of Software: Evolution and Process, 31 (11). e2197-e2197.

DOI: https://doi.org/10.1002/smr.2197

Publisher: Wiley Online Library

Version: Accepted Version

Downloaded from: https://e-space.mmu.ac.uk/624902/

Usage rights: © In Copyright

Additional Information: This is an Author Accepted Manuscript of a paper accepted for publication in Journal of Software: Evolution and Process , published by and copyright Wiley.

Enquiries:
If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

Using blog–like documents to investigate software practice: benefits, challenges and research directions

PREPRINT

Austen Rainer

School of Electronics, Electrical Engineering and Computer Science

Queen's University Belfast, UK

a.rainer@qub.ac.uk

Ashley Williams

Department of Computer Science and Software Engineering

University of Canterbury, NZ

ashley.williams@pg.canterbury.ac.nz

2019

Background: An emerging body of research is using grey literature to investigate software practice. One frequently occurring type of grey literature is the blog post. Whilst there are prospective benefits to using grey literature and blog posts to investigate software practice, there are also concerns about the quality of such material.

Objectives: To identify and describe the benefits and challenges to using blog–like content to investigate software practice, and to scope directions for further research.

Methods: We conduct a review of previous research, mainly within software engineering, to identify benefits, challenges and directions; and use that review to complement our experiences of using blog posts in research.

Results and Conclusion: We identify and organise benefits and challenges of using blog–like documents in software engineering research. We develop a definition of the type of blog–like document that should be of (more) value to software engineering researchers. We identify and scope several directions in which to progress research into and with blog–like documents. We discuss similarities and differences in secondary and primary studies that use blog–like documents, and similarities and differences between the use of blog–

like documents and the use of already established research methods, e.g., interview and survey.

Keywords: social media, blog post, qualitative analysis, case study, survey study, case–survey, systematic review, grey literature, multivocal literature review

Contents

List of Figures

List of Tables

1 Introduction

An emerging body of software engineering researchers (e.g., [26, 25, 27, 75, 23]) argue for, and demonstrate, the benefits of using grey literature to investigate software practice. For example, researchers state that grey literature helps bridge the gap between practice and research, promotes the voice of the practitioner, and helps to better incorporate industrial context. There are however concerns about software engineering research using grey literature, for example concerns about the ability to quality–assure content that is more likely to be subjective and therefore bias. There may also be misconceptions around the status of grey literature, for example whether a grey literature review (GLR) is (un)reasonably treating grey literature with the same evidential value as the primary studies of a Systematic Review (SR).

We are interested in a particular subset of grey literature, a subset we refer to as the *blog–like document*. (We formally define the term *blog–like document* in section 2.) We use the term *blog–like document* because blogs and blog posts are good examples of the kinds of content we are interested in, however the blog post is surprisingly difficult to formally define. Being a subset of grey literature, we are aware that blog–like documents, and their content, inherit both the benefits and challenges of grey literature, e.g., that blog posts help incorporate the practitioners' voice into research etc. The blog–like document may come with its own, additional benefits and challenges, for example the benefit of aggregating and triangulating the views of a particular blogger over time from their multiple blog posts, and the challenge of accommodating the range of feedback often 'attached' to a blog post, e.g., comments, shares, up–votes.

Raulamo–Jurvanen, Mäntylä and Garousi [66], and Soldani, Tamburri and Heuvel [75] report what we believe to be the first systematic grey–literature–only reviews in software engineering. For Soldani *et al.*, 40% of the grey literature used in their study was blog posts. Raulamo–Jurvanen *et al.* and Soldani *et al.*'s work complements the work of Garousi, Felderer and Mäntylä [26, 25, 23, 27] who, in a series of recent papers, promote the need for, value of, and guidelines for multivocal literature reviews (MLRs), which accommodate within them GLRs.

There are other published studies in software engineering research that use blogs, though not in the context of a literature review. For example, Parnin, Treude and Storey [54, 53] have used blogs in their investigations of API documentation. And Pagano and Maalej [50] studied developers' use of project–specific blogs. There has also been work investigating influential bloggers (e.g., [35]) though again not in the context of literature reviews or secondary studies.

The objectives of this paper are:

- to identify and describe benefits to using blog–like documents and con-

tent in software engineering research;

- to identify and describe challenges to using blog–like documents and content in software engineering research; and
- to scope directions for further research to address the challenges and realise the benefits.

The paper makes the following contributions: a set of references, identified through a structured review, concerning the use of blog–like documents; a comparison of the investigation of blog–like documents using literature reviews and case–surveys; a reference definition for blog-like documents; and the identification of five main benefits (and multiple more specific benefits), fourteen main challenges, and fifteen directions for research. We intend for the paper to complement the series of papers by Garousi, Felderer and Mäntylä [26, 25, 23, 27] and the work of Raulamo–Jurvanen, Mäntylä and Garousi [66], and Soldani, Tamburri and Heuvel [75]. The current paper considerably extends a previous paper [64] through, for example, a more formal and more detailed identification and review of relevant literature.

The remainder of the paper is structured as follows: section 2 presents a definition of blog–like documents and uses that definition as the foundation for a structured review of relevant research, including primary and secondary studies in software engineering research. Section 3 reviews and summarises benefits to the use of blog–like documents. Section 4 reviews and summarises challenges to the use of blog–like documents. Section 5 summarises future research directions, contrasts secondary and primary studies, and contrasts information gathered through and from blog–like documents with information gathered through and from interviews and surveys. Finally, in section 6 we provide a brief conclusion, consider threats to the review and identify opportunities to address those threats.

2 Framing the review

2.1 Definition of blog–like documents

In Table 1 we present a reference definition for the use of blog–like documents in software engineering research. We present the definition near the beginning of this paper to help frame the subsequent discussion. The definition emerged from our own experiences of researching blog–like documents together with iteration between search, selection, analysis and reporting of previous literature. The reference definition is based on a set of typical features for blog–like documents; seeks to exclude unsuitable documents, such as those which are irrelevant (based on topic) or of unknown origin (e.g., unknown author); is intended to be used by other software engineering researchers as a reference for developing a definition appropriate to their own research; and is a definition expected to be complemented by a quality–assessment framework (to

be developed by further research) and an appropriate research method (such as an MLR or case–survey).

Table 1: Reference definition of blog–like document in software engineering

#	**A blog–like document typically:**
1	is a publicly accessible document
2	has an identifiable author who is an identifiable software practitioner
3	primarily comprises written content (for pragmatic reasons we focus initially on English) perhaps also with other media (e.g., images, video) and URL links to other content
4	contains personally–written, professionally–oriented content.
5	is published at an online location, i.e. as a webpage.
6	has content that relates to a topic or topics concerning software engineering and its practice
7	is published on a (relatively) frequent basis (typically) in reverse chronological order
8	is (capable of being) revised in response to on–going feedback from readers.
9	is published with a clear date of publication.
10	supports comments and other forms of reader feedback, such as up–votes.

2.2 Rationale for the review

Since 2016 [87], we have been investigating the value and use of blog–like documents as a potential source of information for investigating software practice. From our own empirical work, and through the review of others' empirical work, we have identified potential benefits, recurring challenges, and opportunities and requirements for further research. This paper considerably extends a previous paper [64] to discuss a set of benefits, challenges and research directions drawn from our own experience and complemented by a structured review of prior research.

To reduce bias in our review and to help ensure a balanced review, we performed a series of structured searches for relevant literature, and reviewed the papers identified from those searches.

2.3 Search and selection of literature

Our focus for the review is on primary and secondary studies that empirically investigate software practice using blog–like documents as a clearly identifiable source of data. We intentionally exclude a range of types of social media, such as micro–blogs (e.g., Twitter), videos (e.g., YouTube), instant messaging (e.g., Slack), question–and–answer sites (e.g., Stack Overflow) and email. Excluding these types of social media may appear to over–constrain our search, however we want to focus on the distinctive benefits etc. of

blog–like documents. Each of these types of social media come with their own benefits, challenges and research directions. Some types of social media (notably micro–blogs and question–and–answer sites) have already received considerable attention from research.

We used the ACM Digital Library (DL) as our primary search engine. We chose the ACM DL because of the focus of our review: empirical studies of software practice. We complemented the ACM DL searches with Google Scholar searches. We performed a range of different search queries, the primary query being `<‘‘software engineering’’ [with] ‘‘blog post’’>` for the period 2000 — 2019. The primary query returned 336 results. We reviewed the titles of all articles and also, where appropriate, the abstract and then the full paper. From that review, we identified 42 candidate articles, subsequently reducing this list (through closer inspection of the candidate articles) to 14. We identified a 15^{th} article [54] from our prior experience, and a reviewer of an earlier version of this paper suggested a 16^{th} article. To complement the selected papers on software practice, we used Google Scholar to perform several exploratory searches of the wider literature, selecting three papers as contrasting examples.

Overall, we identified four primary studies, three secondary studies and two methodology papers, all relating explicitly to software practice. To these we added three non–SE papers that were identified through wider searches of the research literature. We complemented those 12 papers with 12 papers from our own research. The 24 papers are summarised in Table 2.

2.4 Edge–case articles that were excluded

We encountered a variety of ‘edge cases’ during our review, and discuss some of these cases as they help to clarify the scope of our review.

While Parnin, Treude and colleagues have papers selected for the review, there are other papers they have published in this area (e.g., [80, 4, 77]) that have been excluded from the review, primarily because those papers are not reporting an empirical study of software practice using blog–like documents. For example, Storey *et al.* [77] describe the nature of blog use amongst developers, but conduct a survey rather than using blog posts directly. Parnin and Treude, with their colleagues Storey and Aniche (and others), are developing a growing body of research relating to practitioners’ use of social media, for which blog–like documents are a subset. Similarly, Garousi, Felderer & Mäntylä (e.g., [23]) are developing a growing body of research incorporating grey literature into software engineering research, for which social media and blog–like documents are subsets. Again, some of Garousi *et al.*’s papers are excluded because the papers are not always reporting an empirical study of software practice using blog–like documents

We frequently found peer–reviewed papers that cited blog posts, in the same way that other peer–reviewed papers are cited. Examples are Rastogi

et al.'s [65] consideration of container debloating, Ernst's [20] consideration of locally accurate prediction models, and Wong and Woepse's [90] industrial survey of *air–gap isolation*. Each of these papers explicitly cite at least one blog post as a reference. Citations to blog posts in these kinds of paper are not being used as literature for a secondary study or as data for a primary study, and therefore were excluded from our study.

We also found a number of papers that discussed the development and evaluation of tools for working with the blogosphere e.g Ferreira *et al.*'s [21] *RetriBlog*, Simões *et al.*'s [74] *WISE Blogs* model, Blanvillain *et al.*'s [8] *BlogForever Crawler*, and Chau *et al.*'s [14] blog mining framework. Lakshmanan and Oberhofer [38] provide a review of approaches and challenges to knowledge discovery in the blogosphere. Such papers were not focused on tools etc. for investigating information about software practice in the blogosphere so, again, these kinds of paper were not formally included in the review. We do however recognise this research in relation to research directions, in section 5.

Another 'edge case' is demonstrated by Taibi *et al.*'s [79] paper on the architectural patterns for microservices[1]. In addition to searching the peer–reviewed literature, Taibi *et al.* conducted forward and backward snowballing searches of the references from their shortlisted peer–reviewed papers. During that snowballing they identified and selected two items of grey literature. They chose these two items because both items occurred frequently in the references of the peer–reviewed papers they'd identified. Formally Taibi *et al.* are not reporting a multivocal literature review or a grey literature review, and neither item of grey literature is explicitly recognised, by Taibi *et al.*, as a blog post. A closer examination of the paper indicates that the authors do not explore the benefits, limitations or research directions of blog–like documents.

2.5 Description of the subsequent analyses of the papers

We identify benefits, challenges and research directions through the combination of our own experience and a review of the independently conducted papers summarised in Table 2. Over an extended period of time, and as a natural activity of our research, each author has read several times each paper. Candidate benefits, challenges and directions therefore emerged over time as we progressed our own research. To ensure a degree of balance and objectivity, we re–read (a final time) the papers identified in the table, annotated each of them, and discussed our annotations, to arrive at a final set of benefits, challenges and research directions.

Our intention is not to develop an exhaustive list of every possible benefit etc. Conversely, aggregating only what others have explicitly reported in

[1]We thank an anonymous reviewer for proposing this paper.

their studies would exclude our own experience, as well as limit what we can identify from others' experience (because authors do not always explicitly state the benefits etc.). We seek to establish a sufficiently coherent body of benefits, challenges and research directions. The clear and coherent statements of the set of benefits, the set of challenges and the set of research directions together provide a reference with which, and against which, future research can progress the use of blog–like documents to investigate software practice.

2.6 The value of blogs for non–software engineering research

Wilson *et al.* [89] report a scoping review of 44 studies in health research to summarise the extent, range, and nature of research activity using blogs in that field of research. Of the 44 studies selected for their review, 38 used blogs for data collection, with 21 of those studies collecting data about experiences, feelings and perceptions, and the remaining 17 collecting data about blogger behaviour. In 11 studies, data was collected from blogs in conjunction with another data source, e.g., interviews, surveys, focus groups. Wilson *et al.*'s scoping study demonstrates the value of blogs in another field of research, and that blogs are frequently used in research for the insights they provide.

Cenite *et al.* [13] investigate bloggers' beliefs and practices in relation to four ethical principles: truth–telling, attribution, accountability and minimizing harm. These beliefs and practices are relevant to the assurance of credibility of blog content. Cenite *et al.* distinguish the personal blog from the non–personal blog. The personal blog is defined as an online diary, maintained by an individual, focusing on content such as daily events and reflections. The non–personal blog is defined to encompass "... all blogs not in the personal blog category; rather than focusing solely on news and commentary, it [the non–personal blog] focuses on areas such as politics, commerce, entertainment or technology." ([13]; p. 578). Cenite *et al.*'s category of the non–personal blog is clearly very broad.

Software engineering researchers would typically not be interested in diaries that report the blogger's *personal* daily events, and her or his reflections on those events. But a software engineering researcher might be interested in some kind of reporting of professional events and the blogger's professional reflections on those events. For example, Bradac *et al.* [10] report the conduct of a process monitoring experiment on software process, in which they used "... log books, personal diaries, and project management notebooks to reconstruct a set of data to represent one developer's experience..." ([10]; p. 781). And Maalej and Happel [45] collected 750,000 work descriptions from three independent and different data sets. Their work descriptions comprised five types: personal notes, time sheets, social media (including blogs), issue tracker comments, and commit messages.

Software engineering researchers are (understandably) concerned about

subjectivity and about the threats to validity that arise with the presence of emotions. Again, there are a number of studies in software engineering (e.g., [30, 49, 91]) that investigate the emotions of software developers. Emotion is explicitly recognised in three of the concepts of *behavioural software engineering* [41]: job satisfaction, self control, and self esteem (with "feeling" explicitly stated in another five concepts).

For Cenite *et al.*'s second category, software engineering researchers would likely be interested in information and commentary, and in bloggers committing to principles of truth–telling, provided such information and commentary related to software engineering. Cenite *et al.* found that bloggers writing non–personal blogs jointly ranked the practices of attribution, truth–telling, and minimizing harm with a mean score of almost 6 on a 7–point Likert scale. In other words, these bloggers self–assessed highly in their ethical beliefs and practices.

In Table 3 we summarise characteristics of bloggers, these characteristics are drawn from two sources: Orbit Media Studios' [17] fourth annual survey of bloggers, and Cenite *et al.*'s [13] study, discussed earlier. The average time to write a blog (over three hours) is approximately three times the length of a usual research interview. We note also the relatively low percentage of females in the non–personal blogging category. This statistic appears consistent with the overall demographics for the software engineering, i.e. a disproportionate number of males work in the sector. Part of the purpose of MLRs and GLRs is to recognise and engage with contrasting perspectives, so as to provide a more balanced understanding of software practice. The relative *percentages* of blog–like documents from males and females is not, in itself, a measure of the *value* of that blog–like content. So while there is a relatively low percentage of females in the non–personal blogging category, the *content* of those non–personal blogs have in principle as much to contribute to our understanding of software practice as the more frequent male perspectives.

2.7 Primary studies using blogs in software engineering research

Parnin, Treude and Storey [54] investigate the motivations and issues of software developers' blogging. They examined blog posts and also surveyed the writers of a set of the blogs they studied. Parnin *et al.* [54] found four main motivators for software bloggers to blog: *personal branding*, *evangelism and recruitment*, *personal knowledge repository*, and *solicit feedback*. The blogger's motivation to blog is an important factor in assessing the credibility of the blog content and, by implication, the quality of the blog content for inclusion in a primary study. Of the four motivators identified by Parnin, Treude and Storey [54], the *personal knowledge repository* is the most relevant to the current paper. Parnin, Treude and Storey [54] write: "The most obvious and frequent use of a blog was to catalog experiences. In our survey,

the most cited benefit from blogging about coding was that the process of writing helped the author learn and remember the information better (93% participants)."

In an earlier study, Parnin and Treude [53] used Google search to survey the extent to which the methods of the jQuery API were documented on the web. Parnin and Treude found that 87.9% of the jQuery methods were covered in blogs. As a contrast, the official API website had a higher coverage (99.4%) and Stack Overflow had a coverage of 84.4%. Parnin and Treude then classified the blog posts by type. The most frequent type of blog post was experience, where "... the post documents development knowledge drawn from a recent experience" [53]. Parnin and Treude quote a blogger who, frustrated with the official jQuery documentation, chose to report their own experience: "I spent over a half hour looking for the best solution to this. Personally, I blame the jQuery documentation. When reading over the jQuery core description it states, starting in version 1.4, that jQuery returns an empty set but offers no method to detect it. Ultimately, I found that .length is the way to go but I wanted to expound on all three methods I discovered."[2]

Pagano and Maalej [50] found that project–specific blog articles (i.e. blog articles about a specific open source project) typically contain fourteen times the word count of version control commit comments, and add value in that they cover high level concepts and functional requirements over the descriptions of low level changes which are typical of commit comments. Pagano and Maalej's analysis of blog content from practitioners within four large, open source communities found that "functional requirements and domain concepts" and "community and contributions" were common topics to all four communities. This leads Pagano and Maalej to conclude that developers blog mainly to promote new features and system requirements, as well as to build up communities around specific topics or technologies.

2.8 Grey–literature studies in software engineering research that use blogs

Raulamo–Jurvanen, Mäntylä and Garousi [66] report what we believe to be the first systematic grey–literature–only review in software engineering research. They identified 60 sources for their review. 59 of the sources reported experiences or opinions, and 7 sources reported examples. It is not clear how many of the 60 sources were blog–like documents though data reported in the paper indicates at least 35 sources had comments, suggesting approximately 60% of the grey literature was blog–like documents (assuming the presence of comments is a defining feature of a blog–like document).

Soldani, Tamburri and Heuvel [75] report one of the first systematic grey–

[2] `http://b-knox.com/181/detect-an-empty-set-in-jquery/`

literature–only reviews in software engineering. Soldani *et al.* selected 51 documents for review, these documents organised into three types: *blog post* (20/51 documents), *whitepaper* (21/51) and *video* (10/51). (Soldani. *et al.*'s paper also refers to a fourth type of document, *industry magazine*, however this type appears to either have been dropped from the analyses or subsumed within the whitepaper type.) Approximately 40% of the grey literature they reviewed were blog posts; and in a befitting coincidence for the current paper, Soldani *et al.* observe that it was with a *blog post* that Lewis and Fowler first introduced the concept of microservices [43].

2.9 Grey literature and white literature in SE research

This paper focuses on blog posts as a particular type of grey literature. In the majority of cases, researchers use the more classical 'white literature' in their research. In terms of Garousi *et al.*'s [23] model of expertise and outlet control, 'white literature' can be understood as that literature where the means and source of both the expertise and the outlet control for that literature are well known. (We revisit Garousi *et al.*'s model in section 4). Because white literature is peer–reviewed there is a 'standardisation' of white literature. For example, white literature often conforms to certain conventions as to the structuring of such literature, has standard citation and referencing styles, and typically uses more formal language. The quality of the content and of the writing tends to be more uniform and of a higher standard (however that standard is defined). There are relatively long processes for preparing, reviewing, revising and publishing such literature. The literature is indexed in central repositories, with meta–data, and each item of literature is indexed with a globally–unique Digital Object Identifier (DOI). But the nature of peer–reviewing often results in publication bias, e.g., where 'negative results' are much less likely to be published.

The 'standardisation' of white literature supports the use of protocols for the systematic review of such literature, such as Systematic Literature Reviews (SLRs; [36]), Systematic Mapping Studies (SMS; [56]), some aspects of Multivocal Literature Reviews (MLRs; [23]), Rapid Reviews [12], and Literature Studies [37]. As a contrast, we have recently developed heuristics to help search grey literature in the absence of such 'standardisation' [62].

Overall, grey literature often lacks the infrastructure and peer–review processes established for white literature and therefore using grey literature in research presents a range of alternative challenges. The current paper identifies and discusses many of these alternative challenges, and proposes directions for research to address those challenges.

3 The benefits of using blog–like documents in SE research

We identify insights that blog–like documents provide into software engineering practice and methodological benefits to the research process. We organise this section according to those two categories.

3.1 Insights into software practice

Blog–like documents are a type of grey literature, and therefore the benefits of studying grey literature also apply to studying blog articles. Garousi *et al.* [24] identify reasons from previous research for utilising grey literature: 1) grey literature provides current perspectives and complements gaps in the formal literature; 2) grey literature may help avoid publication bias (although Garousi et al. acknowledge that the grey literature found may not be representative); and 3) grey literature provides an important perspective on topics.

Software practitioners are often used in research as interviewees and as survey respondents, and are valued for their professional experience, e.g., the propositional and practical knowledge they have gained through practising software engineering. For example, in their survey of 66 practitioners, Procaccino *et al.* write, "Our respondents have experience in software development and, as a result, have opinions based on their professional experience..." ([57], p. 196). Procaccino *et al.* asked the respondents to self–rate their expertise on a Likert scale from *Very inexperienced* to *Expert* and, with a separate measure, Procaccino *et al.* report that their survey respondents experience ranged from less than one year to 37 years.

Bloggers report their experience and that experience ranges from inexperienced bloggers communicating their challenges of maturing in software engineering to highly–experienced, world–renown practitioners communicating their expertise. For research, this experience is often accessed through research methods that engage directly with practitioners, e.g., interviews, surveys, focus groups, and protocol analyses. Blogs and blog articles provide the opportunity to access experience on a much larger scale, albeit retrospectively. In addition to reporting experience, bloggers also write *with* experience. This suggests two dimensions of experience: the degree of experience reported in the blog–like document itself (perhaps as factual stories) and the degree of experience that the bloggers have of software practice in general.

Devanbu *et al.* [19] report on practitioner *beliefs* and Rainer [61] reports on the analyses of practitioners' arguments. Blog–like documents provide the opportunity to investigate a range of practitioner–based concepts and experiences, for example: (empirical) data; practitioners' explanations, e.g., micro–theories for software engineering phenomena; (factual) *stories*, these stories conveying professional experience; practitioners' beliefs about the world; and

the degree to which practitioners use research, e.g., using citations in their blog–like documents.

As a particular example of insights into practice, Parnin and Treude [53] studied the online documentation of APIs. Parnin and Treude analysed 1730 web pages for their coverage of the jQuery API. They identified 376 unique blog posts from the full dataset. They found that the blog posts collectively covered 88% of the API methods. Only the official API covered more (∼99%). Their study suggests that (appropriate) blog–like documents have particular value for documenting APIs. A related value is providing tutorials on the use of APIs, languages etc.

A contrasting example to API documentation is trend analysis. Glance, Hurst and Tomokiyo [28] continuously crawl and analyse blogs to detect trends over time. Trend analysis of blog articles can be a valuable tool for both research and industry.

In addition to trends, blog–like documents can provide early information on new innovations and their adoption. Rogers [68] and Moore [48] both present models of the diffusion of innovation. In both models, there is some concept of innovators and early adopters who are prepared to adopt a new innovation (e.g., a technology) in the absence of objective evidence to support that innovation. Moore [47] distinguished between the early adopters of technology, who are prepared to take greater risks on new technology and tolerate limited evidence of the technologies efficiency, against the majority of adopters, who wait until there is sufficient adoption of technology by others before they will adopt it. This behaviour presents a challenge for research: researchers typically need practitioners to adopt technologies before the researchers can investigate the use and impact of the technology in practice. Blog–like documents provide the prospect for accessing and gathering experiences of innovations from innovators and early adopters. Researchers can then, for example, analyse the content of the blog–like documents, or approach the practitioners for interview, survey etc.

3.2 Methodological benefits

In their scoping review of 44 studies in health research, discussed in section 2, Wilson *et al.* [89] observed that the use of blogs enables researchers to gain instantaneous access to distant populations, provide research clarity and transparency with built–in audit trails, and circumvent the need for lengthy transcriptions of interviews.

Authors of blog–like documents write their documents at different times and also over time, write in different geophysical locations, write about different projects, products and services, post on different platforms, and use different writing software. All these differences suggest independence in the content written.

Much of software engineering is invisible to instrumentation and obser-

vation, and variable too, particularly in the earlier stages of development. Content generated by software practitioners often provides the only viable mechanism for gaining insight into these opaque practices. Blog–like documents about software engineering can potentially provide valuable insights that complement already accepted methods of studying software engineering. For example, Pagano and Maalej [50] found that blog authors tend to explain functional requirements and high level concepts in their blog articles.

Blog–like documents provide the prospect of triangulating and extending practitioner perspectives from different online sites (such as GitHub and Stack Overflow) and also triangulating information drawn retrospectively from online articles with proactively collected information (such as from interviews and surveys). Such triangulation can help to address publication bias.

Garousi, Felderer and Mäntylä [23] present seven questions to help a researcher to decide whether to include grey literature in their multivocal literature review. Their seven questions are based on an earlier set of criteria they developed [25] and on two checklists [6, 2]. The seven questions are presented in Table 4.

3.3 Summary of the benefits of using blogs in research

Table 5 presents a summary of benefits of using blog–like documents, organised into a logical structure. The summary is constructed from a combination of previous work [64], benefits identified earlier in this section, and our own experiences of working with blog–like documents (see Table 2). The benefits are enumerated for reference only; the number does not signify priority. We structure the benefits in terms of: the general (item 1) and specific (item 2) information that blog–like documents can provide, the general value (items 3) that they can provide, the circumstances (item 4) when blog–like documents could be considered for research, and the methodological 'problems' (item 5) that blog–like documents help to address.

4 Challenges of using blog–like documents

We discuss several challenges with blog–like documents, organising these challenges into subsections relating to: definitions and models, frameworks for classifications, quantity and variability of blog–like documents, processes for generating and disseminating blog–like documents, resources, and methodological aspects of blog–like documents.

4.1 Definitions and models

There are many general definitions of grey literature and of blogs. Garden [22], with a paper appositely entitled *Defining blog: A fool's errand or a*

necessary undertaking, reviews the various ways in which the term *blog* has been used within journalism. (She also provides an interesting short history of the term *blog*.) Garden argues that "... the real problem is not that the term blog is difficult to define (it is) but that most scholars are using it in vague, contradictory, ambiguous and imprecise ways... [and] scholars need to provide clear and precise definitions according to the particular research questions asked and the target populations of interest..." ([22]; p. 483). Therefore, there is the challenge to software engineering researchers to be clear about the way they define the term *blog*, and by implication the phrase *blog–like document*, for their research and their research questions. Providing such definitions can help software engineering researchers to define their unit/s of analyses (e.g., the 'type' of blog posts they're considering), help researchers to evaluate the credibility and relevance of those units, and help future systematic reviewers effectively assess the respective primary studies.

Unfortunately, neither Soldani *et al.* [75] nor Garousi *et al.* [23] define blog or blog post for their studies. We therefore provide an indicative list of general definitions of grey literature and blogs in Table 6 (partially extended from [23]). Particularly relevant features of the definitions for the current paper are emboldened in the table. The definitions in Table 6 are not definitions for *blog–like content*, for *blog–like documents*, or for other types of grey literature in *software engineering*. Also, the definitions do not identify a subset of grey literature of more relevance, or of higher–quality, for software engineering research. We therefore provide a reference definition for the term *blog–like document* in section 2.

As well as a lack of definitions of blog–like documents for software engineering, there is also a lack of models, or similar specifications, formally describing the structure and relationships of blog–like documents. Such models would be valuable to researchers to help them build and evaluate repositories of blog–like documents, and build and evaluate tools to crawl and analyses such documents. Such models would also help researchers to evaluate content, an issue we discuss further in subsection 4.4. None of the primary studies, secondary studies, or guideline papers in our review (see Table 2) provide a model of the structure of blog–like documents and their content. Providing such a model is however challenging because of the variability of blog–like documents and their content. To clarify: our intention is not to *prescribe* a model for blog–like documents to which bloggers must conform, but rather to develop conceptual tools that researchers can use to investigate blog–like documents.

4.2 Frameworks for classifying blog–like documents

Garousi, Felderer and Mäntylä [23] present a framework for classifying grey literature. Their framework is reproduced here in Figure 1 together with an alternative that we have developed as an example. Garousi *et al.*'s framework

draws on previous work by Adams *et al.* [2] and Adams *et al.* [1][3]. Later in their paper, Garousi *et al.* use the framework to also help with quality assessment.

An alternative interpretation to Garousi *et al.*'s framework, also presented in Figure 1, plots the three types of grey literature used by Soldani *et al.* [75]. The alternative is intended to illustrate the challenges of classifying grey literature, suggesting that whitepapers, videos and blogs may distribute across the three tiers proposed by Garousi *et al.* [23]. For example, the 20 blog posts included in Soldani *et al.*'s [75] systematic grey literature review would presumably be positioned toward the 1st tier of the framework. (Garousi *et al.* were clear that the three tiers of their framework were not intended to be clearly demarcated, but instead were intended to blend into each other.) As another example, given the variety of experience of different bloggers, one would expect bloggers' blog–like documents to be distributed across the different tiers.

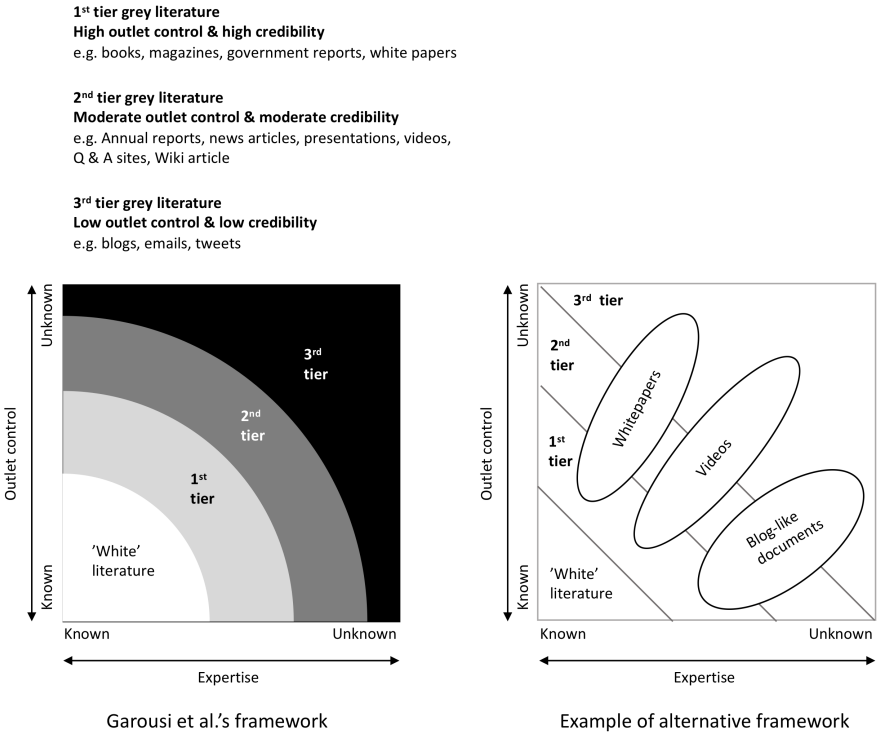

Figure 1: Garousi *et al.*'s framework for classifying grey literature (slightly revised from [23]) with an example alternative framework.

The framework proposed by Garousi *et al.* [23] is useful for appreciating the variation in quality of grey literature. We suggest that a more differentiating

[3]These are two different Adams.

framework for evaluating the quality of blog–like documents is developed. We further suggest that such a framework discriminates aspects of the author/s (e.g., their experience), aspects of process (e.g., the process of generating blog–like documents), aspects of output (e.g., the age of the document vs the content of the document), variation in the content, aspects of the reader feedback, and aspects of the reader; and that such aspects are examined at different levels of analyses.

4.3 The quantity of blog–like documents

There are a very large number of software engineering blogs available on the World Wide Web. For example, almost a decade ago, Lakshmanan and Oberhofer [38] reported that Bansal *et al.*'s *BlogScope* [5], a system for analyzing temporally ordered streaming text online, "... currently tracks more than 36.88 million blogs with 837.39 million posts in the blogosphere... [fetching on average] 14,000 new documents per hour."[4] In terms of software engineering, Choi [15] presents a list of 650 blogs, classified by type (i.e. company, individual/group, and product/technology), and then ordered alphabetically. Panji [51] maintains a curated list of 185 software-related corporate blogs. Merchant [46] maintains a list of over 50 tech blogs. And Abstracta [81] provide a list of 75 blogs and websites on software testing. Being manually curated reference lists, these lists are inevitably much smaller than the full set of blog–like documents written by software engineers on software practice. As a contrast, Soldani *et al.* [75] observes a "... massive proliferation of grey literature [on microservices], with more than 10,000 articles on disparate sub–topics...". The (unknown) quantity of blog–like documents in software engineering presents a challenge to researchers in searching for and selecting the kinds of document that contain content relevant to their research and at a sufficient level of credibility.

4.4 The variability of blog–like documents and content

Soldani *et al.* [75] found it very difficult, in their systematic grey literature review of microservices, to assess the quality of 'their' grey literature, mainly because grey literature lacks a consistent structure.

Blog–like content can vary along a number of dimensions. Indicative dimensions are summarised in Table 7. Of these dimensions, the *Content* dimension is typically the most valuable for research, for it is the *Content* dimension that provides the (most) relevant material for research.

Although the *Content* dimension may be the most valuable to research, each of the dimensions provides challenges for analyses. For example, a common word in blog–like documents for identifying personal experience is (of course) the personal pronoun, "I". Yet the lower case version of that

[4] BlogScope's website, `http://www.blogscope.net/` is no longer responding.

word (letter), "i", is also a common temporary variable name for iterators in programming languages. Most readers would not write that pronoun in lowercase, but a natural language processing pipeline may convert all text to lowercase.

There are no necessarily exclusive relationships between these dimensions. For example, a video might contain an argument about software engineering, an audio file might report a person's experience, a presentation might contain a representation of a real story. All of these may be presented more or less formally, and all may be presented in different natural languages. (As a contrast, the argumentation mining community has focused first on legal corpora, in English, due to its formality and structure [92].) Further, a blog–like document could contain a combination of media and content.

Having analysed a set of blog–like documents, there is then the challenge of concisely presenting the analyses of this variability to others, both for interpretation, but also for traceability of results and reproducibility of findings.

4.5 Processes for generating and publishing blog–like documents

In their study of the motivations and issues of blogging, Parnin, Treude and Storey [54] explain that most developers write up their experiences up to several days after completing the work referred to in the subsequent blog–like document, with only 20% of authors reported writing the blog post the same day as the experience.

The time delay between the experience that inspired the blog–like document and the writing of the document itself potentially impacts the writer's recall of the experience, and therefore raises a threat to the validity of the blog–like content. This time delay, and its affect on recall, also affects other research methods, and possibly has a greater affect on those methods. For example, interviews, surveys and focus groups are all unlikely to take place as close in time and place to the experiences that inform the interview etc.

Also implied in the observations of Parnin and Treude is that a developer writes of a specific experience that inspired that document. And this can, potentially, be an advantage for the use of blog–like content in contrast to other research methods (interviews etc) where respondents may be asked to comment on their general experience and opinions.

The definitions of grey literature presented in Table 6 recognised the lack of a peer–review process and, as a related issue, that grey literature is not formally published. This lack of external mechanism relates to a number of other challenges: the challenge of the quantity of blog–like documents, as there is no mechanism to control such quantity; the challenge of the variability of blog–like documents, as there is no mechanism to control variability; and the challenge of the lack of control of the generation of blog–like documents, as there is no post–generation process to moderate the generation process.

In terms of post–publication guidelines, a brief review of blogs in [64] suggests that common forms of appreciation are comments, shares, and some measure of up–voting or down–voting (or likes, kudos, applause etc). Other measures include follows and re–shares. Similar measures have been used in the analyses of Stack Overflow [18].

A further complication for bloggers can arise with corporate constraints and expectations. Parnin *et al.* [54] observed that, "Several bloggers cited reasons why they might blog in public but not in the workplace. Some reported that corporate blogging might limit exposure and thus usefulness, and also impose censorship. Still, some authors were enthusiastic about blogging at the workplace." ([54], p. 214)

4.6 Ambiguity in the language used in blog posts

Authors of blog posts vary in the formality of language they use in their blog posts, e.g., the language of some blog posts is informal, whilst other blog posts are written as formal, technical articles. The formality of language can vary even within one blog post. As one example of the challenges of language, Swanson *et al.*[78] analysed 50 personal stories drawn from 5,000 posts [29] taken from 44M articles [11]. They used three annotators and achieved an annotator agreement of 0.58. Swanson *et al.* state, "... the annotation task is highly subjective, requiring interpreting the narrative and the author's intention, which prevents us from obtaining high levels of inter–rater agreement." ([78]; p. 175). They also observe earlier in their paper that a previous study [59] found both a high level of annotator agreement and an extremely high machine learning accuracy for Aesop's Fables. Swanson *et al.* [78] infer that the 'classical', written–down stories are therefore easier to work with than blog–posts.

4.7 Resources

4.7.1 Repositories of blog–like documents

Parnin, Treude and Storey [54] state that there exists no central repository of software–related blogs. Choi [15] presents a GitHub repository of 650 software–related blogs (we discussed other GitHub repositories in subsection 4.3) however Choi's repository (like the other repositories) is a manually maintained repository, is not obviously a central repository, and it is not clear how representative Choi's repository is of all software–related blogs. An alternative is *blog aggregators*, such as Planet[5], or news aggregators such as Reddit [4].

In the absence of a central repository of software–related blogs, there remains the challenge of *finding* blogs and their blog posts. As a contrast,

[5] `www.planetplanet.org`

researchers may turn to several central repositories of open source code (e.g., GitHub, Source Forge), and of academic publications. In section 2, we identified several papers reporting tools for finding and working with blogs however none of these tools are targeting blog posts relating to software practice.

Khan *et al.* [35] review (computational) models for identifying influential users in the blogging community. As part of their review, they identify several datasets of blogs and blogs posts. The publicly accessible datasets are summarised in Table 8. None of the datasets considered by Khan *et al.* are specifically relating to software engineering however.

4.7.2 Variation in datasets

Datasets inevitably vary in their content; that is part of their value. But there is the challenge of ensuring that suitably equivalent datasets are gathered for particular analyses. As one example to illustrate the contrasts in the datasets: Pagano and Maalej observed that, "To our surprise, in only 934 of all 50,701 blog posts (1.8%) we found source code paragraphs. On average each of these [source code] posts contained 2.5 code paragraphs." ([50], p. 123). This contrasts with Parnin and Treude ([53]) who observed, "90% of posts [336 posts from 373 posts] had code snippets in the post, a median of 3 code snippets per post" ([53], p. 25). Recall that Pagano and Maalej studied all blog–like documents produced in four open source software projects, whilst Parnin and Treude studied blog–like documents found through Google Searches for (only) the jQuery API method calls. The two studies contrast considerably in the focus of their investigations (e.g., a wide focus on blogging behaviour in software development in contrast to a narrow focus on the use of a particular JavaScript library) and, as a consequence, in the way they searched and selected blog–like documents. As a second example, Aniche and Treude compared the characteristics of a large *r/programming* dataset with a large *Hacker News* dataset. In both datasets they found the majority of blog posts (¿85% in each dataset) are from personal blogs. They also found that the *r/programming* dataset seemed to be more focused on the technical aspects of software engineering, whilst the *Hacker News* dataset had a broader focus on more general topics relating to business, commerce and economics.

4.7.3 Annotated corpora

There are challenges relating to the development of a corpus or corpora of annotated blog articles. A significant challenge is establishing a standard for annotations, e.g., what should be annotated, why and how. Table 9 lists references to corpora of blog–like documents (or similar) that have been marked up for argumentation analyses. A range of other corpora relating to argumentation mining is available at the AIFdb website[6]. We previously

[6] `http://corpora.aifdb.org/`

used a corpora of persuasion essays [76] to validate reasoning indicators [88] to use in Google keyword searches [62].

4.7.4 Proprietary, keyword–based search engines

In section 3, we discussed the value in reporting practitioner experience. There remains the challenge of searching for, selecting, quality–assuring and then synthesising these experiences.

Search engines, such as Google, Bing, and Duck Duck Go, are all keyword–based which means that the searcher has no straightforward way to select the higher–quality documents when performing a search, to select particular types of grey literature, such as blog–like documents, or to select documents that report experience. In addition, all these search engines use proprietary algorithms to index documents online, and also use content delivery networks. These raises challenges around transparency and reproducibility of searches.

The tools available to work with blog–like content, particularly searching for blog–like content, is limited to keyword–based search engines. Unlike academic literature, blog–like content has no specialist search engines, like the ACM DL, IEEEXplore or Google Scholar search engines. As a related issue, blog–like content does not have the editorial processes of *publication* found with academic literature. We distinguish here between the quality–assurance processes, such as academic peer–review, and editorial processes of standardising publications for inclusion in a repository, and the subsequent feedback process.

4.8 Methodology

4.8.1 Guidelines for blog–like documents and content

Parnin, Treude and Storey [54] state that an early research challenge is identifying the best methodology for sampling blogs. Parnin *et al.*, and their colleagues, have not subsequently developed a formal methodology, or guidelines or protocol (to the best of our knowledge), but have instead formulated a design for each study they have conducted. In their study, Parnin *et al.* [54] used keyword–based Google searches to sample blog posts.

In a related paper [53], Parnin and Treude reflect on the limitations to the study they report in that paper. They recognise that online searches can be limited, for example by a mismatch between what the researcher seeks to find with the search terms and what is actually returned by the search engine. Parnin and Treude also propose further research on searching and search terms, e.g., exploring larger sets of search terms, or seeding searches by the search engine's suggestions, or logging search terms used by developers in practice.

Soldani, Tamburri and Heuvel [75] describe the methodology they used for their study. They also recognise the need for a protocol to support the

conduct of systematic grey literature reviews and, as with Parnin *et al.* [54], they developed their own protocol.

Finally, Garousi, Felderer and Mäntylä [23] published a set of guidelines for conducting *multivocal literature reviews* in software engineering, and for including *grey literature* in such reviews. Their guidelines are intended to apply to grey literature in general. But they conclude their paper with the suggestion that future research will, "... refine guidelines for specific types of grey literature sources like *blog articles*..." ([23]; emphasis added here).

4.8.2 Quality assurance

Garousi *et al.* [23] and Soldani *et al.* [75] both develop checklists for the quality–assurance of the documents they studied. Their checklists are summarised in Table 10.

Garousi *et al.* [23] explain that they developed their quality assessment checklist from previous assessment models complemented with their own expertise from previous studies they had conducted. They refer to the "Quality assessment of *sources*... " (emphasis added here) and their source appears to be the item of grey literature being quality–assured. A distinction can be made between the *document* as the source of information, in contrast to the *practitioner* as the source of information which is then documented in a document. More generally, there appears to be distinctions between the source of the document (e.g., an author), the document itself, content within the document, feedback on the document (e.g., readers' comments), and the source of that feedback (from readers).

Soldani *et al.* [75] state that the use of grey literature is "risky" because of the limited amount of rigorous data and analyses in the literature itself. They also found it very difficult to assess the quality of their grey literature mainly because grey literature lacks a consistent structure. As a result, Soldani *et al.* developed a "rudimentary quality control framework" (their term). Soldani *et al.* [75] used a set of inclusion and exclusion criteria, complemented with four additional control factors. They found it was easy/ier to apply their inclusion and exclusion criteria to blog posts and whitepapers, and to subsequently classify them, in contrast to the videos. All these are summarised in Table 10. Soldani *et al.* state in their paper: "An industrial *study* is to be selected if it satisfies all the inclusion criteria, while it is to be excluded if it satisfies at least one of the exclusion criteria... [and] ... we selected only those industrial *studies* that were satisfying four additional control factors..." ([75]; emphasis added here)

4.9 Summary of challenges

Table 11 summarises the main challenges identified in the preceding subsections.

5 Directions for further research

5.1 Overview

Given the challenges identified in section 4, there are a number of directions for further research. We summarise these research directions in Table 12 indicating progress made on each direction. In the following subsections, we discuss two particular areas for further research: literature reviews compared to case–surveys, and blog–like documents compared to already–established methods of information collection from practitioners.

5.2 Literature reviews and case–surveys

In subsection 4.8 we observed that there exist guidelines for the conduct of secondary studies (i.e. protocols for MLRs and GLRs, though not specifically of blog–like documents) but no guidelines for the conduct of primary studies that use blog–like documents as a data source. There is the implication of course that blog–like documents can be used in both secondary studies and primary studies.

Instead of investigating blog–documents in terms of literature and its review (cf. [23, 75]) the researcher could investigate blog–like documents in terms of *data* and their analyses. Each blog–like document could constitute a case (cf. [70, 71]) comprising one or more units of analyses (e.g., textual content relating to the topic or topics relevant to the research). In their primary studies of blog–like documents, Parnin, Treude, Storey and Aniche [4, 54, 53] have not treated blog–like documents as literature. Neither have they explicitly treated blog–like documents as cases for analyses. By contrast, Pagano and Maalej's [50] automated analyses of blog–like documents, written in relation to the development of four open source projects, explicitly treats such documents as data. As a final example, one way to interpret Rainer's [61] investigation of defeasible reasoning is as the analysis of one case (one blog–like document written by Joel Spolsky) with multiple units of analyses (multiple instances of reasoning that comprised argumentation, citations to sources, and reports of experience as stories).

Because of the quantity of blog–like documents available and the need to conduct search and selection of cases, we recognise the need to complement the depth–focus of case studies with a breadth–search. Surveys are an obvious method to conduct such breadth–search and –selection. A research approach that combines case study and survey study is already available: the case–survey, e.g., [34, 55]. We are developing a variant of the case–survey methodology intended specifically for blog–like documents. A preliminary version was reported in [64]. Table 13 presents an overview to the case–survey method.

An natural question to ask is which of the two approaches, secondary studies (and therefore literature reviews) or primary studies (e.g., case–surveys)

is the 'better' approach; or, alternatively, under what circumstances one approach is more appropriate than the other. The questions presented in Table 4 appear to apply equally to either approach. Table 14 provides a comparison between secondary studies and primary studies of blog–like documents. Two distinguishing features between secondary study and primary study appear to be: first, how the researcher conceives of blog–like documents (as literature or as data); second, the extent of coverage of documents sought by the researcher. Systematic Reviews *ideally* seek comprehensive (if not complete) coverage of primary studies. But such an objective is extremely difficult to obtain with grey literature and blog–like documents (because of the challenges recognised in this paper). Hence, a representative sample of blog–like documents may be a more pragmatic approach. It then becomes important to *sample* blog–like documents, though sampling from a population of blog–like documents is also extremely difficult for the population is hard to define, as is a sampling frame. Nevertheless, a primary study that employs the case–survey design more explicitly recognises the sampling of documents.

5.3 Collecting information from practitioners

Software engineering researchers use a range of research methods for collecting information from software practitioners, the most common methods being interview and questionnaire survey. For example, Seaman [73] writes that interviews are often used to collect historical data from the memories of interviewees.

A similarity across interview, questionnaire survey and blog–like document is that the practitioner provides information from their experience and the formation of their beliefs. In other words, there is no inherent difference between the three methods in how the information reported in interview, questionnaire survey and blog–like document is initially gathered, of formed, by the practitioner. The inherent challenges of a practitioner recalling information about historical events when writing a blog–like document are similar to the challenges of a practitioner recalling such information for an interview; or indeed for a survey.

There are, however, differences between interview, questionnaire survey and blog–like document in *how* the researcher gathers information from the practitioner. Two differences are, first, the degree of control available to the researcher in gathering information; and, second, the degree of interactivity available between practitioner and researcher for gathering information. For interview, the researcher has a relatively high degree of control in choosing who to interview, and in choosing the questions to ask, including follow–up questions; and the researcher retains a relatively high degree of interactivity through the opportunity to immediately ask questions. This high degree of interactivity provides the opportunity for the researcher to both help the

practitioner recall information, and also to help assess the validity of the information as it is being recalled.

For questionnaire survey, the researcher has again a relatively high degree of control, but a significantly reduced opportunity for interactivity. For blog–like documents, the researcher has some degree of control in the choice of which blog–like documents to work with, but no opportunity to ask clarifying questions (of that blog–like document) and no opportunity for interactivity. For both questionnaire survey and blog–like documents, there is limited opportunity to help the practitioner recall information. Reviewing a series of blog–like documents from one author could, in some circumstances, provide the opportunity to clarify the experiences and beliefs of the respective author.

We have noted, in section 2.7, that practitioners tend to write their blog–like documents within a few days of the experiences that form the content of the blog–like document. This suggest an immediacy of information for blog–like documents that often may not be available for interviews. Such immediacy does not of course entirely address threats to the recall and reporting of information.

Unlike interviews and surveys, blog–like documents provide for a greater opportunity for the public scrutiny and subsequent revision of the blog–like documents. For example, a typical research interview might last about an hour and it is usually only within that hour that the researcher has the opportunity to clarify the information provided by the interviewee. (Some investigations provide the opportunity for the interviewee to review a copy of the subsequent transcript or notes of the interview.) The 'document' from the interview is then fixed. By contrast, the blog–like document allows for reader feedback over an unconstrained time period and provides the opportunity for the writer to revise the document over that time period too (though, of course, that opportunity may not always or even often be taken). In addition, a blog as a series of blog–like documents provides the opportunity for the researcher to investigate the consistency and coherence of the practitioner's experiences and beliefs.

From the above discussion, we draw two main observations: first, each type of data and each research method provide advantages and disadvantages to the researcher, with the implication that a multi–source and multi–method approach to research is likely to be more effective and credible. Second, this discussion re–emphasises the importance of establishing the quality (the validity) of the particular blog–like documents that are selected and analysed in the research, just as a researcher should establish the quality (credibility) of a respondent to a survey or an interviewee.

6 Conclusion

In this paper, we focus on the use of blog–like documents to investigate software practice. We propose a reference definition of the blog–like document. We identify and discuss the (prospective) benefits to using blog–like documents in research, and the challenges of using such documents. We also identify research directions and indicate progress on each of those directions. We provide a core set of references on the use of blog–like documents in primary and secondary studies into software practice. We contrast the conception of blog–like documents as literature for review with the conception of blog–like documents as data to be analysed. We also contrast the collection of blog–like documents, as a method of gathering data, with the conduct of interviews and surveys: all three methods depend on practitioners reporting their experience of software practice. We do not claim that our lists of benefits, challenges and directions are exhaustive however we do believe they represent a substantive and coherent body of insights with which, and against which, future research can progress the use of blog–like documents to investigate software practice.

There are, inevitably, threats to the validity of this review. One threat is the small number of primary and secondary studies identified. This is partly a consequence of the number of such studies in software engineering that explicitly use blog–like documents to investigate software practice, and partly a consequence of our search and selection process. We might have widened the review to consider other types of social media, however our objective was to concentrate on the distinctive benefits etc. of blog–like documents. We briefly discussed other types of social media in section 2. One natural development to the current paper is to review these other types and to compare the resulting benefits etc. A second appropriate development would be to conduct a review of primary and secondary studies in other subject areas, e.g., to more comprehensively review papers in healthcare (like [89]). It is of course unlikely that such papers would relate to investigating software practice.

A second threat relates to the processes we have used to conduct our searches, to select papers, and to analyse the resulting papers. We intentionally focused on ACM DL searches, complemented by Google Scholar searches. There is the possibility that we missed relevant papers by not using other bibliographic search engines. For example, one of the four primary studies we selected was not returned in the ACM DL searches. For the particular 'missed study', we would expect this paper to be indexed in the ACM DL. There is also the possibility of bias, for although both authors independently reviewed the papers, we have worked closely together on many papers, and so there may be a bias common to both authors. An obvious way to examine this bias is for others to independently analyse the papers identified in Table 2 as well as to independently perform searches to find and select papers.

As well as the conduct of further reviews of published studies, there is also the opportunity to gather information about benefits, challenges etc. directly from researchers. We have previously conducted a survey of researchers' opinions of the credibility of blog posts [86] though this survey does not address benefits etc. As indicated by the number of studies identified in the review, one potential difficulty with a survey or interviews is the limited number of researchers who have direct experience of using blog–like documents to investigate software practice.

Finally, conducting further primary and secondary studies that use blog–like documents as data or as literature would produce a more substantive body of research against which the community can make a more informed judgement on the value and challenges of using blog–like documents to investigate software practice.

acknowledgements

We thank Professor Claes Wohlin for his suggestion to use the case survey method. We thank the reviewers and editor for their many thoughtful and constructive comments.

References

[1] Jean Adams, Frances C Hillier-Brown, Helen J Moore, Amelia A Lake, Vera Araujo-Soares, Martin White, and Carolyn Summerbell. "Searching and synthesising 'grey literature'and 'grey information'in public health: critical reflections on three case studies". In: *Systematic reviews* 5.1 (2016), p. 164.

[2] Richard J Adams, Palie Smart, and Anne Sigismund Huff. "Shades of grey: guidelines for working with the grey literature in systematic reviews for management and organizational studies". In: *International Journal of Management Reviews* 19.4 (2017), pp. 432–454.

[3] Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim. "A Benchmark Dataset for Automatic Detection of Claims and Evidence in the Context of Controversial Topics". In: *Proceedings of the First Workshop on Argumentation Mining*. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 64–68. URL: `http://www.aclweb.org/anthology/W/W14/W14-2109`.

[4] Maurıcio Aniche, Christoph Treude, Igor Steinmacher, Igor Wiese, Gustavo Pinto, Margaret-Anne Storey, and Marco Aurélio Gerosa. "How modern news aggregators help development communities shape and share knowledge". In: *IEEE/ACM 40th International Conference on Software Engineering (ICSE'18)*. IEEE. 2018, pp. 499–510.

[5] Nilesh Bansal, Fei Chiang, Nick Koudas, and Frank Wm Tompa. "Seeking stable clusters in the blogosphere". In: *Proceedings of the 33rd international conference on very large data bases, University of Vienna, Austria*. Sept. 2007, pp. 806–817. URL: `http://http://www.vldb.org/conf/2007/`.

[6] Karen M Benzies, Shahirose Premji, K Alix Hayden, and Karen Serrett. "State-of-the-evidence reviews: advantages and challenges of including grey literature". In: *Worldviews on Evidence-Based Nursing* 3.2 (2006), pp. 55–61.

[7] Or Biran and Owen Rambow. "Identifying justifications in written dialogs". In: *5th IEEE International Conference on Semantic Computing (ICSC)*. IEEE. Palo Alto, California, USA, September 18-21, 2011, pp. 162–168.

[8] Olivier Blanvillain, Nikos Kasioumis, and Vangelis Banos. "Blogforever crawler: techniques and algorithms to harvest modern weblogs". In: *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS'14)*. ACM. Thessaloniki, Greece, 2–4 June, 2014, p. 7.

[9] Filip Boltužić and Jan Šnajder. "Back up your stance: Recognizing arguments in online discussions". In: *Proceedings of the First Workshop on Argumentation Mining*. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 49–58. URL: `http://www.aclweb.org/anthology/W/W14/W14-2109`.

[10] Mark G Bradac, Dewayne E Perry, and Lawrence G Votta. "Prototyping a process monitoring experiment". In: *IEEE Transactions on Software Engineering* 20.10 (1994), pp. 774–784.

[11] Kevin Burton, Akshay Java, Ian Soboroff, et al. "The ICWSM 2009 spinn3r dataset". In: *Third Annual Conference on Weblogs and Social Media (ICWSM 2009)*. San Jose, California, 17-–20 May: The AAAI Press, Menlo Park, California, 2009.

[12] Bruno Cartaxo, Gustavo Pinto, and Sergio Soares. "The Role of Rapid Reviews in Supporting Decision-Making in Software Engineering Practice." In: *22nd International Conference on Evaluation and Assessment in Software Engineering (EASE'18)*. University of Canterbury, Christchurch, New Zealand, 28–29 June, 2018, pp. 24–34.

[13] Mark Cenite, Benjamin H Detenber, Andy WK Koh, Alvin LH Lim, and Ng Ee Soon. "Doing the right thing online: a survey of bloggers' ethical beliefs and practices". In: *New Media & Society* 11.4 (2009), pp. 575–597.

[14] Michael Chau, Jennifer Xu, Jinwei Cao, Porsche Lam, and Boby Shiu. "A blog mining framework". In: *IT Professional* 11.1 (2009), pp. 36–41.

[15] Kilim Choi. *Software Engineering Blogs*. URL: `https://github.com/kilimchoi/engineering-blogs#-individuals`. (accessed: 04.08.2018).

[16] Michael Conniff. "Just what is a blog, anyway". In: *Online Journalism Review* 29 (Sept. 2005). URL: `http://www.ojr.org/p050929/`.

[17] Andy Crestodina. *Blogging Statistics and Trends: The 2017 Survey of 1000+ Bloggers*. URL: `https://www.orbitmedia.com/blog/blogging-statistics/`. (accessed: 07.11.2018).

[18] Daniel Hasan Dalip, Marcos André Gonçalves, Marco Cristo, and Pavel Calado. "Exploiting user feedback to learn to rank answers in q&a forums: a case study with stack overflow". In: *Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval*. ACM. Dublin, Ireland, 28 July – 1 August, 2013, pp. 543–552.

[19] Premkumar Devanbu, Thomas Zimmermann, and Christian Bird. "Belief & evidence in empirical software engineering". In: *IEEE/ACM 38th International Conference on Software Engineering (ICSE'16)*. IEEE. Austin, Texas, USA, 14–22 May, 2016, pp. 108–119.

[20] Neil A. Ernst. "Bayesian Hierarchical Modelling for Tailoring Metric Thresholds". In: *Proceedings of the 15th International Conference on Mining Software Repositories*. MSR '18. Gothenburg, Sweden: ACM, 2018, pp. 587–591. ISBN: 978-1-4503-5716-6. DOI: `10.1145/3196398.3196443`. URL: `http://doi.acm.org/10.1145/3196398.3196443`.

[21] Rafael Ferreira, Rinaldo Lima, Jean Melo, Evandro Costa, Fred Freitas, and Henrique Pacca. "RetriBlog: a framework for creating blog crawlers". In: *27th Annual ACM Symposium on Applied Computing*. ACM. Trento, Italy, 26–30 March, 2012, pp. 696–701.

[22] Mary Garden. "Defining blog: A fool's errand or a necessary undertaking". In: *Journalism* 13.4 (2012), pp. 483–499. DOI: `10.1177/1464884911421700`. eprint: `https://doi.org/10.1177/1464884911421700`. URL: `https://doi.org/10.1177/1464884911421700`.

[23] Vahid Garousi, Michael Felderer, and Mika V Mäntylä. "Guidelines for including grey literature and conducting multivocal literature reviews in software engineering". In: *Information and Software Technology* 106 (2019), pp. 101–121.

[24] Vahid Garousi, Michael Felderer, and Mika V Mäntylä. "Guidelines for including the grey literature and conducting multivocal literature reviews in software engineering". In: *arXiv preprint arXiv:1707.02553* (2017).

[25] Vahid Garousi, Michael Felderer, and Mika V Mäntylä. "The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature". In: *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering*. ACM. Limerick, Ireland, 1–3 June, 2016, p. 26.

[26] Vahid Garousi and Mika V Mäntylä. "A systematic literature review of literature reviews in software testing". In: *Information and Software Technology* 80 (2016), pp. 195–216.

[27] Vahid Garousi and Mika V Mäntylä. "When and what to automate in software testing? A multi-vocal literature review". In: *Information and Software Technology* 76 (2016), pp. 92–117.

[28] Natalie Glance, Matthew Hurst, and Takashi Tomokiyo. "Blogpulse: Automated trend discovery for weblogs". In: *WWW 2004 workshop on the weblogging ecosystem: Aggregation, analysis and dynamics*. Vol. 2004. New York. 2004.

[29] Andrew Gordon and Reid Swanson. "Identifying personal stories in millions of weblog entries". In: *Third International Conference on Weblogs and Social Media, Data Challenge Workshop, San Jose, CA*. Vol. 46. 2009.

[30] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. "Happy software developers solve problems better: psychological measurements in empirical software engineering". In: *PeerJ* 2 (2014), e289.

[31] Ivan Habernal, Judith Eckle-Kohler, and Iryna Gurevych. "Argumentation Mining on the Web from Information Seeking Perspective." In: *Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing (ArgNLP)*. Forlì–Cesena, Italy, 21–25 July, 2014.

[32] Susan C Herring, Lois Ann Scheidt, Sabrina Bonus, and Elijah Wright. "Bridging the gap: A genre analysis of weblogs". In: *37th annual Hawaii international conference on system sciences*. IEEE. Big Island, Hawaii, USA, 5–8 January, 2004, 11–pp.

[33] Kentaro Inui, Shuya Abe, Kazuo Hara, Hiraku Morita, Chitose Sao, Megumi Eguchi, Asuka Sumida, Koji Murakami, and Suguru Matsuyoshi. "Experience mining: Building a large-scale database of personal experiences and opinions from web documents". In: *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent*

Technology-Volume 01. IEEE Computer Society. Washington, DC, USA, 9–12 December, 2008, pp. 314–321.

[34] Marlen C Jurisch, Petra Wolf, and Helmut Krcmar. "Using the case survey method for synthesizing case study evidence in information systems research". In: *Proceedings of the Nineteenth Americas Conference on Information Systems*. Chicago, Illinois, USA, 15–17 August, 2013.

[35] Hikmat Ullah Khan, Ali Daud, Umer Ishfaq, Tehmina Amjad, Naif Aljohani, Rabeeh Ayyaz Abbasi, and Jalal S Alowibdi. "Modelling to identify influential bloggers in the blogosphere: A survey". In: *Computers in Human Behavior* 68 (2017), pp. 64–82.

[36] B. Kitchenham and S Charters. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. 2007.

[37] Marco Kuhrmann, Daniel Méndez Fernández, and Maya Daneva. "On the pragmatic design of literature studies in software engineering: an experience-based guideline". In: *Empirical software engineering* 22.6 (2017), pp. 2852–2891.

[38] Geetika Lakshmanan and Martin Oberhofer. "Knowledge discovery in the blogosphere: Approaches and challenges". In: *IEEE Internet Computing* 14.2 (2010), pp. 24–32.

[39] Amanda Lawrence, John Houghton, Julian Thomas, and Paul Weldon. "Where is the evidence? Realising the value of grey literature for public policy and practice". In: (2012). DOI: 10.4225/50/5580B1E02DAF9. URL: https://researchbank.rmit.edu.au/view/rmit:39652.

[40] Carol Lefebvre, Eric Manheimer, and Julie Glanville. "Searching for studies". In: *Cochrane handbook for systematic reviews of interventions: Cochrane book series* (2008), pp. 95–150.

[41] Per Lenberg, Robert Feldt, and Lars Göran Wallgren. "Behavioral software engineering: A definition and systematic literature review". In: *Journal of Systems and software* 107 (2015), pp. 15–37.

[42] Len L. Levin. "Literature Search Strategy Week: Len Levin on Understanding and Finding Grey Literature". In: (2014). URL: https://escholarship.umassmed.edu/lib_articles/171.

[43] James Lewis and Martin Fowler. *Microservices: A Definition of this New Architectural Term*. 2014. URL: https:%20//www.martinfowler.com/articles/microservices.html.

[44] Marco Lippi and Paolo Torroni. "Argumentation mining: State of the art and emerging trends". In: *ACM Transactions on Internet Technology (TOIT)* 16.2 (2016), p. 10.

[45] Walid Maalej and Hans-Jorg Happel. "From work to word: How do software developers describe their work?" In: *6th IEEE International Working Conference on Mining Software Repositories (MSR'09)*. IEEE. Vancouver, Canada, 16–17 May, 2009, pp. 121–130. URL: http://2009.msrconf.org/.

[46] Amit Merchant. *Tech Blogs*. URL: https://github.com/amitmerchant1990/tech-blogs. (accessed: 08.02.2019).

[47] GA Moore. *Crossing the Chasm: Marketing and Selling Technology Project. Revised Edition*. 2009.

[48] Geoffrey A Moore. "Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers (Collins Business Essentials)". In: *HarperBusiness, New York* (1991).

[49] Alessandro Murgia, Parastou Tourani, Bram Adams, and Marco Ortu. "Do developers feel emotions? an exploratory analysis of emotions in software artifacts". In: *11th working conference on mining software repositories*. ACM. Hyderabad, India, 31 May – 1 June, 2014, pp. 262–271. URL: http://2014.msrconf.org/.

[50] Dennis Pagano and Walid Maalej. "How do developers blog?: an exploratory study". In: *8th working conference on Mining software repositories*. ACM. Honolulu, Hawaii, USA, 21–22 May, 2011, pp. 123–132. URL: http://2011.msrconf.org/.

[51] Muhammad Panji. *Engineering Blogs*. URL: https://github.com/sumodirjo/engineering-blogs. (accessed: 08.02.2019).

[52] Joonsuk Park and Claire Cardie. "Identifying appropriate support for propositions in online user comments". In: *Proceedings of the First Workshop on Argumentation Mining*. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 29–38. URL: http://www.aclweb.org/anthology/W/W14/W14-2101.

[53] Chris Parnin and Christoph Treude. "Measuring API documentation on the web". In: *Proceedings of the 2nd international workshop on Web 2.0 for software engineering*. ACM. 2011, pp. 25–30.

[54] Chris Parnin, Christoph Treude, and Margaret-Anne Storey. "Blogging developer knowledge: Motivations, challenges, and future directions". In: *IEEE 21st International Conference on Program Comprehension (ICPC)*. IEEE. San Francisco, California, USA, 20-21 May, 2013, pp. 211–214.

[55] Kai Petersen, Deepika Badampudi, Syed Muhammad Ali Shah, Krzysztof Wnuk, Tony Gorschek, Efi Papatheocharous, Jakob Axelsson, Severine Sentilles, Ivica Crnkovic, and Antonio Cicchetti. "Choosing Component Origins for Software Intensive Systems: In-House, COTS, OSS or

Outsourcing?—A Case Survey". In: *IEEE Transactions on Software Engineering* 44.3 (2018), pp. 237–261.

[56] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. "Guidelines for conducting systematic mapping studies in software engineering: An update". In: *Information and Software Technology* 64 (2015), pp. 1–18.

[57] J Drew Procaccino, June M Verner, Katherine M Shelfer, and David Gefen. "What do software practitioners really think about project success: an exploratory study". In: *Journal of Systems and Software* 78.2 (2005), pp. 194–203.

[58] Arun Qamra, Belle Tseng, and Edward Y Chang. "Mining blog stories using community-based and temporal clustering". In: *Proceedings of the 15th ACM international conference on Information and knowledge management*. ACM. 2006, pp. 58–67.

[59] Elahe Rahimtoroghi, Reid Swanson, Marilyn A Walker, and Thomas Corcoran. "Evaluation, orientation, and action in interactive storytelling". In: *Proc. of Intelligent Narrative Technologies* 6 (2013).

[60] Austen Rainer. "The practitioner as information–provider to software engineering research". In: *Submitted for publication* (2019).

[61] Austen Rainer. "Using argumentation theory to analyse software practitioners' defeasible evidence, inference and belief". In: *Information and Software Technology* 87 (2017), pp. 62–80.

[62] Austen Rainer and Ashley Williams. "Heuristics for improving the rigour and relevance of grey literature searches for software engineering research". In: *Information and Software Technology* 106 (2019), pp. 231–233.

[63] Austen Rainer and Ashley Williams. "Mining professional experience in grey literature: an exploratory review". In: *submitted for publication*.

[64] Austen Rainer and Ashley Williams. "Using blog articles in software engineering research: benefits, challenges and case-survey method". In: *25th Australasian Software Engineering Conference (ASWEC 2018)*. Adelaide, Australia, 26–30 November, 2018. URL: `https://aswec2018.net/`.

[65] Vaibhav Rastogi, Chaitra Niddodi, Sibin Mohan, and Somesh Jha. "New directions for container debloating". In: *Proceedings of the 2017 Workshop on Forming an Ecosystem Around Software Transformation*. ACM. 2017, pp. 51–56.

[66] Päivi Raulamo-Jurvanen, Mika Mäntylä, and Vahid Garousi. "Choosing the Right Test Automation Tool: A Grey Literature Review of Practitioner Sources". In: *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*. EASE'17. Karlskrona, Sweden: ACM, 2017, pp. 21–30. ISBN: 978-1-4503-4804-1. DOI: 10.1145/3084226.3084252. URL: http://doi.acm.org/10.1145/3084226.3084252.

[67] Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M Khapra, Ehud Aharoni, and Noam Slonim. "Show me your evidence-an automatic method for context dependent evidence detection". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 2015, pp. 440–450.

[68] Everett M Rogers. *The diffusion of innovation 5th edition*. 2003.

[69] Sara Rosenthal and Kathleen McKeown. "Detecting opinionated claims in online discussions". In: *6th IEEE International Conference on Semantic Computing (ICSC)*. IEEE. Palermo, Italy, 19–21 September, 2012, pp. 30–37.

[70] Per Runeson and Martin Höst. "Guidelines for conducting and reporting case study research in software engineering". In: *Empirical software engineering* 14.2 (2009), p. 131.

[71] Per Runeson, Martin Höst, Austen Rainer, and Bjorn Regnell. *Case study research in software engineering: Guidelines and examples*. John Wiley & Sons, 2012.

[72] J. Schöpfel and D.J. Farace. "Grey literature". In: *Encyclopedia of Library and Information Sciences* (2010). cited By 37, pp. 2029–2039.

[73] Carolyn B Seaman. "Qualitative methods in empirical studies of software engineering". In: *IEEE Transactions on software engineering* 4 (1999), pp. 557–572.

[74] João Simões, Frederico Azeiteiro, and Jorge Bernardino. "WISE Blogs: A Special Blog Search Engine". In: *8th International C* Conference on Computer Science & Software Engineering*. ACM. Yokohama, Japan, 13–15 July, 2015, pp. 144–145.

[75] Jacopo Soldani, Damian Andrew Tamburri, and Willem-Jan Van Den Heuvel. "The pains and gains of microservices: A Systematic grey literature review". In: *Journal of Systems and Software* 146 (2018), pp. 215–232.

[76] Christian Stab and Iryna Gurevych. "Annotating argument components and relations in persuasive essays". In: *25th International Conference on Computational Linguistics: Technical Papers (COLING'14)*. Dublin, Ireland, August, 2014, pp. 1501–1510.

[77] Margaret-Anne Storey, Leif Singer, Brendan Cleary, Fernando Figueira Filho, and Alexey Zagalsky. "The (r) evolution of social media in software engineering". In: *Future of Software Engineering (FOSE'14)*. ACM. Hyderabad, India, 31 May – 7 June, 2014, pp. 100–116.

[78] Reid Swanson, Elahe Rahimtoroghi, Thomas Corcoran, and Marilyn Walker. "Identifying narrative clause types in personal stories". In: *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*. 2014, pp. 171–180.

[79] Davide Taibi, Valentina Lenarduzzi, and Claus Pahl. "Architectural patterns for microservices: a systematic mapping study". In: *8th International Conference on Cloud Computing and Services Science (CLOSER'18)*. Funchal, Madeira, Portugal, 19–21 March, 2018, pp. 221–232. URL: http://closer.scitevents.org/?y=2018.

[80] Christoph Treude and Margaret-Anne Storey. "Effective communication of software development knowledge through community portals". In: *19th ACM SIGSOFT symposium and the 13th European conference on Foundations of software engineering*. ACM. Szeged, Hungary, 5–9 September, 2011, pp. 91–101.

[81] Kalei White. *75 Best Software Testing Blogs*. URL: https://abstracta.us/blog/75-best-software-testing-blogs/. (accessed: 04.08.2018).

[82] Ashley Williams. "Assessing the credibility of online articles: an extended abstract". In: *Emerging Researchers' Forum, co–located with 22nd International Conference on Evaluation and Assessment in Software Engineering (EASE'18)*. University of Canterbury, Christchurch, New Zealand, 28–29 June, 2018.

[83] Ashley Williams. "Do software engineering practitioners cite research on software testing in their online articles?: A preliminary survey." In: *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*. ACM. University of Canterbury, New Zealand, 28–29 June, 2018, pp. 151–156.

[84] Ashley Williams. "Using reasoning markers to select the more rigorous software practitioners' online content when searching for grey literature". In: *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*. ACM. University of Canterbury, New Zealand, 28–29 June, 2018, pp. 46–56.

[85] Ashley Williams and Austen Rainer. "Do software engineering practitioners cite research in their online articles? A larger scale replication". In: *23rd International Conference on Evaluation and Assessment in Software Engineering (EASE'19)*. Copenhagen, Denmark, 2019.

[86] Ashley Williams and Austen Rainer. "How do software engineering researchers assess the credibility of practitioner–generated blog posts?" In: *23rd International Conference on Evaluation and Assessment in Software Engineering (EASE'19)*. Copenhagen, Denmark, 2019.

[87] Ashley Williams and Austen Rainer. "Identifying practitioners' arguments and evidence in blogs: insights from a pilot study". In: *Software Engineering Conference (APSEC), 2016 23rd Asia-Pacific*. IEEE. 2016, pp. 345–348.

[88] Ashley Williams and Austen Rainer. "Toward the use of blog articles as a source of evidence for software engineering research". In: *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*. ACM. 2017, pp. 280–285.

[89] Elena Wilson, Amanda Kenny, and Virginia Dickson-Swift. "Using blogs as a qualitative health research tool: a scoping review". In: *International journal of qualitative methods* 14.5 (2015), p. 1609406915618049.

[90] Sunny Wong and Anne Woepse. "Software Development Challenges with Air-gap Isolation". In: *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. ESEC/FSE 2018. Lake Buena Vista, FL, USA: ACM, 2018, pp. 815–820. ISBN: 978-1-4503-5573-5. DOI: 10.1145/3236024.3275526. URL: http://doi.acm.org/10.1145/3236024.3275526.

[91] Michal R Wrobel. "Emotions in the software development process". In: *6th International Conference on Human System Interaction (HSI)*. IEEE. Gdansk, Poland, 6–8 June, 2013, pp. 518–523. URL: http://hsi.wsiz.rzeszow.pl/.

[92] Adam Wyner, Raquel Mochales-Palau, Marie-Francine Moens, and David Milward. "Approaches to text mining arguments from legal cases". In: *Semantic processing of legal texts*. Springer, 2010, pp. 60–79.

Table 2: Summary of previous research that informs the current paper

Authors	**Year**	**Title**
Guidelines for MLRs in software engineering research		
Garousi, Felderer & Mäntylä [23]	2019	Guidelines for including grey literature and conducting multivocal literature reviews in software engineering
Garousi, Felderer & Mäntylä [25]	2016	The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature
Secondary studies (MLRs and GLRs) of software practice using blog–like documents as sources		
Soldani, Tamburri & Heuvel [75]	2018	The pains and gains of microservices: a systematic grey literature review
Raulamo–Jurvanen, Mäntylä and Garousi [66]	2017	Choosing the right test automation tool: a grey literature review of practitioner sources
Garousi & Mäntylä [27]	2016	When and what to automate in software testing? A multi-vocal literature review
Primary studies of software practice using blog–like documents as data		
Aniche *et al.* [4]	2018	Where does Google find API documentation?
Parnin, Treude & Storey [54]	2013	Blogging developer knowledge: Motivations, challenges, and future directions
Parnin & Treude [53]	2011	Measuring API Documentation on the Web
Pagano and Maalej [50]	2011	How do developers blog?: an exploratory study
Examples from wider research: Journalism and *health research*		
Wilson *et al.* [89]	2015	Using blogs as a qualitative health research tool: a scoping review
Garden [22]	2012	Defining blog: A fool's errand or a necessary undertaking
Cenite et al. [13]	2009	Doing the right thing online: a survey of bloggers' ethical beliefs and practices
Our research on blog–like documents		
Rainer [60]	—	The practitioner as information–provider to software engineering research.
Rainer & Williams [63]	—	Mining professional experience in grey literature: an exploratory review
Williams & Rainer [85]	2019	Do software engineering practitioners cite research in their online articles? A larger scale replication
Williams & Rainer [86]	2019	How do software engineering researchers assess the credibility of practitioner–generated blog posts?
Williams [82]	2018	Assessing the credibility of online articles: an extended abstract
Williams [84]	2018	Using reasoning markers to select the more rigorous software practitioners' online content when searching for grey literature
Williams [83]	2018	Do software engineering practitioners cite research on software testing in their online articles? A preliminary survey.
Rainer & Williams [62]	2018	A method for improving the rigour and relevance of grey literature searches for software engineering research
Rainer & Williams [64]	2018	Using blog articles in software engineering research: benefits, challenges and case–survey method.
Rainer [61]	2017	Using argumentation theory to analyse software practitioners' defeasible evidence, inference and belief
Williams & Rainer [88]	2017	Toward the use of blog articles as a source of evidence for software engineering research.
Williams & Rainer [87]	2016	Identifying practitioners' arguments and evidence in blogs: insights from a pilot study.

Table 3: Characteristics of blogging and bloggers

Characteristic	**Explanation**
Attributes from the fourth annual Orbit Media Studios blogger survey [17] (N=1377)	
Average time to write	On average it takes 3hrs 20mins to write a blog post
Frequency of blogging	Bloggers typically write several times a month
Average length of post	The average length of a post is 1142 words.
Updating previous posts	55% of bloggers update posts at least sometimes.
Statistics from Cenite et al.'s study [13] for their non–personal bloggers only (N=332)	
Mean age (SD)	34.9 (12.2) years
Female bloggers (%)	19.2%
Reason/s for blogging	36% Provide commentary
	21% Provide information
	11% Express thoughts and feelings
Primary audience	48% People not known to me personally

Table 4: Questions to decide whether to include grey literature in a review (phrasing is modified slightly from [23])

#	**Criterion**
1	Is the subject 'complex' and not solvable by considering only the formal literature?
2	Is there a lack of volume or quality of evidence, or a lack of consensus of outcomes measurement in the formal literature?
3	Is the contextual information important to the subject under study?
4	Is the goal of the review to validate or corroborate scientific outcomes with practical experiences?
5	Is the goal of the review to challenge assumptions or falsify results from practice using academic research, or vice versa?
6	Would a synthesis of insights and evidence from the practical and research communities be useful to either or both of those communities?
7	Is there a large volume of practitioner sources indicating high practitioner interest in the topic?

Table 5: Summary of benefits of using blog–like documents in research

#	**Benefit**
1	In general, blog–like documents : • provide information on practitioners' contemporary perspectives on important topics relevant to practice and to research; and • promote the voice of the practitioner.
2	In particular, blog–like documents provide (access to) information on the practitioner's: • experience and inexperience of theirs' and others' software practice; • motivations for that practice; • values relating to that practice; • emotions relating to that practice; • beliefs about software practice; • empirical data from their practice; and • explanations of that practice.
3	In providing such information, blog–like documents: • help bridge the divide between research and practice; • complement the research literature by 'filling in gaps' in research; and • help to counteract bias findings, as a result of publication bias in the research literature.
4	Blog–like documents should be considered when: • the topic of the research is complex; • the topic is not 'solvable' by using only the peer–reviewed research literature; • there is a lack of quantity and/or quality of best evidence from research, or a lack of consensus in the research; • context is important to the study of the topic; • the researcher intends to challenge existing assumptions and findings, either in research or practice, or both; • a synthesis of practice and research would be valuable to either or both communities; • the researcher intends to consider trends over time; and • the researcher seeks to better understand, assess or demostrate the impact of research in relation to a particular topic.
5	Methodologically, the use of blog–documents in research helps researchers to: • assess and address publication bias; • compensate for the (un)availability of other sources of evidence; • increase research visibility into actual software practice; • access harder–to–access practitioners, e.g., due to logistics, or demographics; • gather information for the research in a non–invasive way; • scale–up their research to, or with, larger samples; • complement and triangulate with, other sources of data • provide an audit trail of their research, e.g., reduced need for preparing transcripts of interviews. • replicate each others' study through public access to original data; and • circumvent the need for lengthy and expensive transcriptions

Table 6: General definitions of grey literature and blogs, with emboldened emphasis on particularly relevant features

#	**Definition of grey literature**	**Date**	**Source**
1	"... grey literature is composed of knowledge artefacts that are **not the product of peer-review processes** characterizing publication in scientific journals..." [emphasis added here]	2014	[39]
2	"... anything that has not been published in a traditional format or, in library parlance, **lacks bibliographic control**, meaning it can be hard to look up. This includes things such as conference proceedings, conference posters, dissertations and theses, government/institutional reports and raw data... luckily, much of it is now online ... 'Institutional Repositories'... Government agencies – federal, state, provincial, etc. ... generate many reports that contain excellent data... [B]logs, Tweets or Facebook postings... can also be a great place to locate valuable information not found elsewhere." [emphasis added here]	2014	[42]
3	"[grey literature]... is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers, i.e., where publishing is not the primary activity of the producing body"	2011	[72]
4	"... literature that is **not formally published** in sources such as books or journal articles... " [emphasis added here]	2008	[40]
#	**Definition of blog**	**Date**	**Source**
5	"There is no need to define 'blog'. A blog is merely a tool that lets you do anything from change the world to share your shopping list. I resist even calling it a medium; it is a means of sharing information and also of interacting: It's more about conversation than content ... Blogs are whatever they want to be. Blogs are whatever we make them. Defining 'blog' is a fool's errand."	2005	[16]
6	"... **frequently modified** Web pages in which **dated entries** are listed in reverse **chronological sequence**."	2004	[32]

Table 7: Indicative dimensions of variability in blog–like documents

Dimension	**Explanation and examples**
Quality of written language	For example the formality of language.
Natural language	Most research appears to focus on English but there are of course a very wide range of other languages to consider.
Media	Video
	Text
	Image
	Sound
	Presentations
'Encoding' of the media	Text with, for example, HTML
	(Proprietary) binary formats, e.g., Adobe PDF
Structure	Headings, sub–headings
Content	Reasoning, e.g., claims, reasons, arguments
	Opinions
	Reporting of actual experience, perhaps as a 'war story'
	Code–related information, e.g., source code, documentation, API
	Web links, e.g., URLs
	(Tables of) data
	Citations

Table 8: Datasets of blog posts (selected from Khan *et al.*'s list [35])

Dataset	**Comment**
	`URL for site`
	`URL for download`

TUAW	A dataset from the Unofficial Apple Weblog dedicated to Apple products and services (TUAW is now obsolete but the dataset is available).
	`http://tuaw.com`
	`http://socialcomputing.asu.edu/uploads/1251628491/TUAW-dataset.zip`

Digg	A social news website
	`http://www.digg.com`
	`http://socialcomputing.asu.edu/uploads/1296588940/Digg-dataset.zip`

BlogCatalog	A social blog directory that manages blogs and bloggers.
	`http://www.blogcatalog.com/`
	`http://socialcomputing.asu.edu/uploads/1252092625/BlogCatalog-dataset.zip`

Engadget	A multilingual weblog providing information on digital electronics.
	`http://http//www.http://www.engadget.com/`
	`http://users.sch.gr/lakritid/code.php?c=3`

Everything2	A blogging community spanning a diversity of content, e.g., poetry, fiction, humour.
	`https://everything2.com`
	`https://everything2.com/title/dataset`

Table 9: Corpora of blog–like documents (or similar) used in argumentation mining and experience mining

Reference	**Date**	**Document type**
Argumentation mining corpora (selected from [44])		
Rinott *et al.* [67]	2015	Wikipedia pages
Aharoni *et al.* [3]	2014	Wikipedia pages
Boltuzic and Snajder [9]	2014	User comments
Habernal *et al.* [31]	2014	Web documents
Park and Cardie [52]	2014	User comments
Rosenthal and McKeown [69]	2012	Blogs, forums
Biran and Rambow [7]	2011	Blog threads
Experience mining corpora (drawn from a currently unpublished literature view [63])		
Swanson *et al.*[78]	2014	229 blogs containing personal stories
Inui *et al.* [33]	2008	One year's worth of Japanese weblog posts
Qamra *et al.* [58]	2006	Over one million blog posts collected from crawling the blogosphere from December 2004 through September 2005

Table 10: Garousi *et al.*'s [23] and Soldani *et al.*'s [75] quality checklists

Garousi *et al.*'s [23] quality checklist items	
Authority of the producer	
Is the publishing organization reputable?	Is an individual author associated with a reputable organization?
Has the author published other work in the field?	Does the author have expertise in the area?
Methodology	
Does the source have a clearly stated aim?	Does the source have a stated methodology?
Is the source supported by authoritative, contemporary references?	Are any limits clearly stated?
Does the work cover a specific question?	Does the work refer to a particular population or case?
Objectivity	
Does the work seem to be balanced in presentation?	Is the statement in the sources as objective as possible? Or, is the statement a subjective opinion?
Is there vested interest?	Are the conclusions supported by the data?
Date: Does the item have a clearly stated date?	
Position regarding related resources: Have key related grey literature or formal sources been linked to / c	
Novelty	
Does it enrich or add something unique to the research?	Does it strengthen or refute a current position?
Impact: A normalisation of several impact metrics	
Number of citations	Number of backlinks
Number of media shares	Number of comments
Number of views	
Outlet type: 1st Tier, 2nd Tier, or 3rd Tier (see Figure 1)	

Soldani *et al.*'s [75] quality checklist items	
Inclusion criteria	
I1 The study discusses the industrial application of microservices.	I2 The study discusses the benefits or shortcomings of microservice design, development or operation.
I3 The study reports on direct experiences, opinions or practices on microservices by educated practitioners.	I4 The study refers to a practical case-study of design, development or operation of microservices.
Exclusion criteria	
E1 The study does not offer details on design or implementation of microservices.	E2 The study is not referred to industrial cases or other factual evidence.
E3 The benefits or pitfalls of microservices are not justified/quantified by the study.	E4 The study does not provide scope and limitations of proposed solutions/patterns.
E5 The study does not offer evidence of a practitioner perspective.	
Additional control factors	
C1 Practical experience: A study is to be selected only if it is written by practitioners with 5+ experience in service-oriented design, development and operation, or if it refers to established microservices solutions with 2+ years of operation.	C2 Industrial case-study: A study is to be selected only if it refers to at least 1 industrial case-study where a quantifiable number of microservices are operated.
C3 Heterogeneity: The selected studies reflect at least 5 top industrial domains and markets where microservices were success-	C4 Implementation quantity: The selected studies refer to/show implementation details for the benefits and pitfall they discuss, so

Table 11: Summary of challenges to using blog–like documents and content in research, grouped by theme

#	**Challenge**
Foundations, e.g., There are a lack of...	
1	Formal definitions of grey literature and blog–like documents and content;
2	Formal models of blog–like documents and content, in particular; • a data model of blog–documents and content; and • a process model of the creation, review and publication of blog–documents and content;
3	Discriminating frameworks for evaluating the quality of blog–like documents and content, and classifying those documents and content;
The inherent nature of blog–like documents and content, e.g., There are challenges managing...	
4	The very large quantity of blog–like documents
5	The variability of blog–like content and documents
6	The uncertain process for generating, publishing and revising the content of blog–like documents;
7	The ambiguity of language in the content of blog posts;
Resources, e.g., There are a lack of...	
8	Repositories of blog–like documents;
9	Tools to work with blog–like documents and content, for example: • to select the higher–quality documents when performing a search; and • to select particular types of blog–like documents, e.g., those reporting experience, values, explanations etc.
10	Datasets and annotated corpora of blog–like documents, including; a lack of 'standards' for describing and comparing datasets; and a lack of 'standards' for annotation of such datasets;
Search engines	
11	Search engines with proprietary indexing algorithms (and content delivery networks) introduce challenges for the independent reproducibility of search results.
Quality–assurance, e.g. there is a lack of:	
12	Well–developed and accepted checklist for the quality–assurance of various aspects of blog–like documents including: • the author; • the document; • the content of the document, e.g., claims; • the readers' assessment of the credibility of the document; • the readers; • the readers' feedback on the document, e.g., comments, shares, upvotes;
Methodology	
13	The evidential value of blog–like content;
14	The appropriate research methods to use with blog–like documents and content.

Table 12: Summary of research directions

#	**Research direction**
Foundations	
1	Replicate the work of Cenite et al. [13] (see section 2) to better understand the ethical beliefs and practices of software practitioners who write blog–like documents.
2	Validate and evaluate the reference definition.
3	Develop and validate a data model, particularly for the different kinds of content (see Table 7).
4	Develop and validate a process model of the writing and revision of blog–like documents. A preliminary model has been developed [64] and subsequently revised [60]
5	Develop a more refined framework for classifying blog–like documents (see section 4 and Figure 1).
The inherent nature of blog–like documents	
6	Develop and evaluate guidance for conducting searches. (We have already developed search heuristics [62].)
7	Develop and evaluate search tools. We have developed a suite of tools collectively known as COAST and are evaluating that suite of tools (`https://github.com/zedrem/coast_core` and `https://github.com/zedrem/coast_search`). In section 2 we identified a number of existing studies of searching, crawling and mining the blogosphere, e.g., Ferreira *et al.* [21], Simões *et al.* [74], Blanvillain *et al* [8], Chau *et al.* [14], and Lakshmanan & Oberhofer [38].
8	Investigate processes that bloggers use to generate, review and edit, publish and post–publish revise their blog–like documents
Resources	
9	Identify, classify and evaluate datasets and corpora. Table 8 and Table 9 provide initial examples. A further set is provided in [63].
10	Develop guidelines for describing datasets to help researchers aggregate or otherwise assemble datasets for use in their research
11	Develop guidelines for annotating corpora to help researchers use those corporate in subsequent research.
Guidelines and checklists...	
12	Develop and evaluate guidelines specifically for the use of blog–like documents in software engineering research. Such guidelines may distinguish between the use of blog–like documents in a primary study compared to a secondary study. Garousi *et al.*'s [23] guidelines for MLRs are a resource from which to start to develop guidelines specifically for blog–like documents.
13	Develop and evaluate a quality assessment checklist specifically for the use of blog–like documents in software engineering research. Garousi *et al.*'s [23] checklist and Soldani *et al.*'s [75] checklist (see Table 10) are two resources from which to start to develop checklists specifically for blog–like documents.
14	Investigate the blogging characteristics of blogging behaviour (cf. section 2 and particularly 3) e.g., how long practitioners take to write a blog etc. The most obvious method of investigation would be the survey (or interview) however this requires self–assessment. An alternative could be to collect statistics from a blogging platform. These characteristics can then be compared with other datesets, e.g., those listed in Table 9.
Methodology	
15	Systematically compare the framing of blog–like documents as *cases* in contrast to blog–like documents as *literature*, and consider the implications for the evidential value of blog–like documents and for research methodology

Table 13: Summary of case–survey protocol

#	**Stage**
S1	Establish research objectives, requirements and rationale
	Establish rationale for the research. Establish research objectives. Establish general research questions and specific sub–questions, prioritise those, and structure them. Define propositions and/or hypotheses (if any). Define variables. Assure the quality of the research objectives etc.
S2	Define and source cases.
	Define the case (unit of analysis). Design the search strategy, search queries and search terms. Ensure consistent alignment of cases, search strategy etc and research questions. Execute searches. Download the search results. Perform backward snowballing on URL links in the downloaded search results. Perform post–search quality filtering. Perform quality assurance checks on the downloaded data and the filtered data.
S3	Define the survey of cases.
	Identify more specific exploratory research questions. Identify variables. Operationalise variables.
S4	Extract data from the surveyed cases.
S5	Analyse extracted data.
	Test the propositions and hypotheses. Answer the research questions.
S6	Disseminate the findings of the study.

Table 14: Similarities and differences between secondary studies and primary studies of blog–like documents

#	**Similarities between secondary studies and primary studies**
1	Secondary studies and primary studies both distinguish the blog–like document from the content of that document;
2	Secondary studies and primary studies both investigate naturally–written texts that have varying degrees of formality and written quality;
3	Secondary studies and primary studies both seek 'coverage' of the body of literature;
4	Secondary studies and primary studies both (potentially) deal with a large volume of literature, i.e. have scaling challenges;
5	Secondary studies and primary studies both seek to apply quality–criteria to select the more relevant and higher–quality documents (though the quality–criteria may differ, perhaps significantly);
6	Secondary studies and primary studies both use similar keyword–based search engines to search for appropriate literature (though there are differences in the search capabilities of the respective search engines, and in the repositories available);
7	Secondary studies and primary studies both analyse documents with unknown or uncertain (editorial) processes of document–generation;
8	Secondary studies and primary studies both analyse documents with unknown or uncertain processes of knowledge–generation or –accumulation;
#	**Differences between secondary studies and primary studies**
9	Secondary studies and primary studies differ in their conception of the object of study:
	1. Secondary studies review publications; whilst
	2. Primary studies analyse data;
10	Secondary studies and primary studies may differ in their degree of coverage:
	1. Secondary studies seek a (relatively) comprehensive coverage of the literature; whilst
	2. Primary studies seek a representative sample of the population of literature;