eISSN : 2287-4577 pISSN : 2287-9099
http://www.jistap.org
Vol. 7 No. 2 June 30, 2019

Journal of Information Science Theory and Practice

JISTaP

KISTI www.kisti.re.kr

Indexed/Covered by SCOPUS, LISA, DOAJ, and CrossRef

General Information

Aims and Scope

The *Journal of Information Science Theory and Practice (JISTaP)* is an international journal that aims at publishing original studies, review papers and brief communications on information science theory and practice. The journal provides an international forum for practical as well as theoretical research in the interdisciplinary areas of information science, such as information processing and management, knowledge organization, scholarly communication and bibliometrics. JISTaP will be published quarterly, issued on the 30th of March, June, September, and December. JISTaP is indexed in the Scopus, Korea Science Citation Index (KSCI) and KoreaScience by the Korea Institute of Science and Technology Information (KISTI) as well as CrossRef. The full text of this journal is available on the website at http://www.jistap.org

Indexed/Covered by

Publisher

Korea Institute of Science and Technology Information
66, Hoegi-ro, Dongdaemun-gu, Seoul, Republic of Korea
(T) +82-2-3299-6102
(F) +82-2-3299-6067
E-mail: jistap@kisti.re.kr
URL: http://www.jistap.org

Managing Editor: Ji-Young Kim, Eungi Kim

Copy Editor: Ken Eckert

Design & Printing Company: SEUNGLIM D&C

4F, 15, Mareunnae-ro, Jung-gu, Seoul, Republic of Korea
(T) +82-2-2271-2581~2
(F) +82-2-2268-2927
E-mail: sdnc@sdnc.co.kr

Open Access and Creative Commons License Statement

This paper meets the requirements of KS X ISO 9706, ISO 9706-1994 and ANSI/NISO Z39.48-1992 (Permanence of Paper)

Table of Contents

JISTaP

Vol. 7 No. 2 June 30, 2019
Journal of Information Science Theory and Practice • http://www.jistap.org

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice
eISSN : 2287-4577 pISSN : 2287-9099

Research Paper
J Inf Sci Theory Pract 7(2): 06-22, 2019
https://doi.org/10.1633/JISTaP.2019.7.2.1

Information Systems in Interdisciplinary Research: Analytic and Holistic Ways to Access Information Science Knowledge

Volkmar P. Engerer*

Department of Information Studies, University of Copenhagen,
Copenhagen, Denmark
E-mail: volkmar.engerer@hum.ku.dk

ABSTRACT

The paper explores how information science knowledge can be used systematically in digital, interdisciplinary research settings and gives a conceptual analysis of the relationship between information science knowledge as donor and other research as receiver in an interdisciplinary project environment. The validity of the approach is demonstrated by the author's work on the project "The Primacy of Tense: A. N. Prior Now and Then." The study proposes a hybrid approach, combining analysis and synthesis. The analytical component identifies information systems, assigns an information system type to them, and accesses the information science knowledge associated with that type. The synthetic part focuses on the connections between information systems according to the receiver discipline's practices. The paper makes explicit the actions of experienced information professionals, thereby making their expertise accessible to others. The analytical and synthetic strategies are explained by linking them to two modes of researchers in the receiver discipline, how they act as researchers and what they know about it. The paper offers information professionals concrete assistance with identification of the appropriate strategy for accessing professional knowledge and taking appropriate actions and development decisions.

Keywords: interdisciplinarity, project collaboration, information system, research communication, digital research environment, Arthur N. Prior

Open Access

Received date: April 09, 2019
Accepted date: May 08, 2019

***Corresponding Author:** Volkmar P. Engerer
Associate Professor
Department of Information Studies, University of Copenhagen, Søndre Campus, build. 4, Njalsgade 76, DK-2300 Copenhagen S, Denmark
E-mail: volkmar.engerer@hum.ku.dk

1. INTRODUCTION

How can information science inform research communication and, in particular, the development of systems for research communication? This question arose from the author's engagement in a research project with which he is currently affiliated. The funded, Denmark-based research project "The Primacy of Tense: A. N. Prior Now and Then" (Prior Project Group, 2017) involves researchers interested in the New Zealand philosopher and logician Arthur Norman Prior, collaborating with information scientists affiliated with the Department of Information Studies, University of Copenhagen. The main tasks of the information science group, to which the author of this paper belongs, include development of Danish Prior websites associated with the project in order to enhance communication and collaboration between Prior researchers throughout the world, and to make Prior's unpublished manuscripts accessible in transcribed, digitised form. In 2017 the focus was on modernising the Danish Prior websites, accelerating the Prior Virtual Lab's production of transcribed manuscripts and making them more accessible via the Internet. The project group has reported elsewhere on some of the information science issues raised by the project (Engerer, Roued-Cunliffe, Albretsen, & Hasle, 2017; Engerer & Sabir, 2018) and more visions for the development of the digital Prior resources along information science lines can be found in Engerer and Albretsen (2017). This paper draws frequently on the author's practical work in the project to illustrate its main points.

Online research communication and collaboration (and their scientific study) are relatively recent phenomena and are strongly connected to the rise of the networked personal computer and the World Wide Web (Tredinnick, 2007). Internet-based systems for research communication include, for example, research portals (Becker et al., 2012), digital platforms for scientific collaboration ('collaboratories,' cf. Finholt, 2002; Olson et al., 2008) and, more recently, 'cyber-infrastructures' in e-science (Borgman, 2007; Elsayed, Madey, & Brezany, 2011). All of these types of system are well-researched interdisciplinary objects that are explored by researchers with very diverse research interests and theoretical backgrounds. For example, socio-constructivist learning theory and the concept of co-evolution from Luhmann's system theory have been combined in order to shed light on learning and knowledge building in online communities (Kimmerle, Cress, & Moskaliuk, 2012; Notari & Honegger, 2012) and complexity theory has been used to model research teams as complex systems interacting on several levels (Vasileiadou, 2012).

Thus from a science-sociological perspective, it does not seem surprising that applied and technological perspectives dominate research on communication and collaboration in research. Typical practical endeavours include research into taxonomies and types of research collaboration infrastructures (Bos et al., 2008), lists of success criteria for online collaboration (Olson et al., 2008), designs for evaluation procedures for collaboration projects (Ramage, 2010), and research into related issues such as ways of managing interdisciplinary digital communication and collaboration (Cummings & Kiesler, 2008). More technological issues are grid computing, big science, data mining and dataspace (Elsayed, Madey, & Brezany, 2011; Finholt, 2002), coding, standards and mark-up techniques (Eggert, 2009; Flanders, 2012), digital collaboration tools (Zaugg, West, Tateishi, & Randall, 2011), and others.

Sometimes it is somewhat unclear how the results of these various strands of research connect with each other and what consequences they have for a broader and more general perspective on digital research communication and collaboration; nevertheless this is a promising and exciting interdisciplinary field of real substance. Investigations into research communication can improve our understanding of how researchers interact with technology, with other researchers and the public, and with information—often all at the same time. The proposals presented in this paper concern work in an interdisciplinary project setting, in which the integration of information science knowledge into the development of digital research communication and collaboration systems plays a crucial role.

Some of the studies cited draw on specialised, sometimes fragmentary, theoretical frameworks, whereas others prefer more coherent, discipline-specific approaches to research communication and collaboration involving 'packages' of knowledge accrued by a discipline over the course of its history. Examples of this latter strategy include the use of a system of interconnected psychological concepts and theories such as 'impersonality' or 'being one's self' in a psychological analysis of blogging (Gurak & Antonijevic, 2012; for a more general account see Wallace, 2001) and a discussion of 'cyber-ethnography,' which redefines sociological inquiry and traditional ethnographic methodology (field work, participant observation, and text-as-data) for the new online environments (Robinson & Schulz, 2012).

It is not easy to find discipline-based studies of information science that deal specifically with research collaboration communication and research into it. It is true that much work in this field appeals to the importance and ubiquity of information, information behaviour, and other related informational concepts in relation to researchers' learning, collaboration, and research

practices;[1] yet despite the numerous references to informational concepts, scholars only occasionally address information science knowledge directly.[2] This suggests a discrepancy between the widespread acceptance of the relevance of information science concepts to research communication and the application of disciplinary knowledge from the field of information science by those studying research communication.

This paper attempts to address this gap and so it presents some methodological and theoretical insights which should be useful both in research on scholarly collaboration and to those in the field of information science who are supporting the development of research collaborations. In a support setting information science is not expected to make a direct, disciplinary contribution to answering a project's research questions (hereinafter referred to as the 'domain,' 'domain research,' etc.). Instead, it contributes as a 'support discipline,' being used to explore and enhance the digital resources of a research project in another discipline, the domain. On the other hand, the 'constitutive' role of information science in interdisciplinary collaboration means that as a disciplinary field information science contributes to the project's research on the same footing as the other project disciplines (for example philosophy, history, or logic) (Engerer & Sabir, 2018). This paper is exclusively concerned with information science in a supporting role.

2. RELATING INFORMATION SCIENCE TO OTHER RESEARCH

To clarify how information science links up with other research in a project environment, in our case the 'logical/philosophical' component of the Prior project, information science knowledge is related to three aspects of domain research:

a) conceptual knowledge in the domain (e.g., logic, philosophy)
b) research activities in the domain (e.g., discussing a logical/philosophical argument, applying for funding) and
c) practical knowledge about these domain research activities (b) that is held by domain researchers (e.g., logicians, philosophers).

These three related aspects can be interpreted as distinct forms of domain knowledge, conceptual and practical (in the cases of [a] and [c]) (Krohn, 2010), and types of communicative and non-communicative activity carried out by a professional agent in an academic domain for intellectual or coordination purposes (in the case of [b]) (this distinction is explained in more detail below) (Erkens, Prangsma, & Jaspers, 2006). The semantic relationship between these two dimensions, information science and domain research, is one of 'transfer,' where the academic 'donor' is information science and the academic 'recipient' the research domain.

The first part of this section describes the three knowledge/activity components a)-c) in more detail. The second part discusses three ways in which information science can be related to each of the three aspects of domain knowledge/activity.

2.1. The Domain Knowledge/Activity Dimension

On the most basic level 'conceptual domain knowledge' (a) is internalised knowledge of a discipline's research objects (logical entities and concepts, philosophical arguments, proofs, etc.), as found in, for example, books, articles or other media from the domain (Goldsmith, Johnson, & Acton, 1991).

'Domain activities' (b) are the communicative and non-communicative acts ('doings,' cf. Schatzki, 1996) of logicians and philosophers within their domain. These activities can serve intellectual or coordinating functions. Where they serve an intellectual function they are directed at joint knowledge building (Cress & Kimmerle, 2008) and addressing the research questions of the domain or project, and are therefore closely related to conceptual domain-specific knowledge (a). The coordinating functions include all the practical, research-related questions in a domain that require communication, for example, project coordination, research dissemination, funding and cooperation over publishing projects, and organisation of meetings and conferences (Rolland & Potter, 2017). Coordinating labour is only indirectly linked to research questions and domain knowledge, and there is a clear hierarchy of research activities, with intellectual labour regarded as proper

[1] Informational concepts referred to in collaborative research include information needs, accessibility of information, and access points of collaboration platforms (Borgman, 2007, p. 2; Elsayed et al., 2011, p. 270), questions of content and mark-up in digital information and websites (Eggert, 2009, p. 75), the idea of information as a shared, accessible, and created commodity in knowledge collaboration (Kimmerle et al., 2012), the buzzword 'information overload' (Cummings & Kiesler, 2008, p. 113), digital libraries for research (Finholt, 2002, p. 79), and Borgman's notion of 'information infrastructure,' which emphasises the information/data dichotomy in relation to modern research collaborations (Borgman, 2007, ch. 3). Furthermore, references to the importance of tacit and presupposed knowledge in digital communication (Finholt, 2002, p. 96) and the conceptual value of distinguishing between 'information' and 'knowledge' when studying research communication (information is easier to mediate than knowledge) (Bos et al., 2008, p. 54) are often supported by citing information science theory.

[2] Cf. Hockey who emphasises the positive role that 'information specialists' (a kind of practising information scientist) play in collaborations with researchers on digital humanities projects, but does not directly refer to information science sources (Hockey, 2012, p. 87). An exception is Christine Borgman, who is exploring in great detail how information science concepts can be used to understand digital research communication and collaboration, a topic she treats in her book *Scholarship in the digital age* (Borgman, 2007).

research and coordinating labour as merely facilitative.

'Domain community knowledge' (c) highlights that a domain researcher is part of a disciplinary community of researchers, with a collective history, norms, quality standards, criteria for good arguments and good research, and academic motivational systems (Elsayed et al., 2011; Tompkins, Perry, & Lippincott, 1998). This set of norms, criteria, and standards guides researchers' activities in the domain (b), whether intellectual (e.g., discussion about a research paper) or coordinating (e.g., department meetings, discussing funding possibilities, etc.).

Domain-specific procedural knowledge typically consists of sequences of activities which 'make sense' (Dervin, Foreman-Wernet, & Lauterbach, 2003) in the context of the domain concerned. Thus, domain community knowledge determines the order in which tasks are executed and the nature of communicative activities in the context of the domain concerned. One example would be the structuring of processes a domain uses for information seeking (Case, 2012; Engerer & Gudiksen, 2016), from seeking a reference and accessing the full text document to checking its relevance to the researcher's interests, downloading the reference into the researcher's reference-managing program and relating it to other references therein, downloading the full-text document to a target destination with accepted ordering principles (for example a list in a Dropbox folder that is ordered alphabetically by author), and so on. This also demonstrates that not all activities in the domain are necessarily communicative; an activity such as downloading a document makes sense for the agent alone by virtue of being preceded by saving the corresponding reference (a prerequisite for subsequent citation) and succeeded by the naming of the document file according to the ordering principle of the researcher's private repository (to ensure it is easily retrievable) (Østerlund, Snyder, Sawyer, Sharma, & Willis, 2015). These three aspects of domain research are linked as Fig. 1.

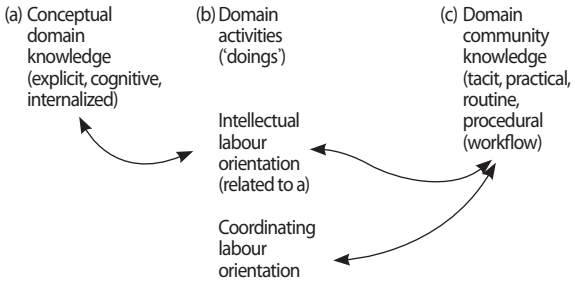


Fig. 1. Relationships between conceptual domain knowledge (a), activities in the domain (b), and a community's tacit, practical knowledge how to do things in the domain (c).

A community's tacit, practical domain knowledge (c) organises the activity within the domain (b) and structures activities having directly to do with the domain's disciplinary objects of knowledge (intellectual labour) and other more practical research-related communications (coordinating labour). At the same time, these communicative activities instantiate and externalise the tacit, procedural knowledge of the research community (c). There is also a two-way relationship between intellectual labour-related communication and cognitive knowledge-building (a). This connection highlights that research is not an isolated, individual process of cognitive knowledge-building, but, like all learning (Wenger, 1998), is a social and communicative activity as well.

2.2. Information Science Perspectives on the Domain

These three knowledge/activity aspects of researchers in a domain yield three distinct perspectives on information science knowledge and how it is brought into play in an interdisciplinary project where information science is the supporting discipline.

Perspective 1, which relates to conceptual domain knowledge (a), entails interdisciplinary interaction in which information science concepts such as 'information' and 'knowledge' are integrated into logical and philosophical reasoning and research (Floridi, 2011). Today interdisciplinarity is a research discipline in its own right (Frodeman, Klein, & Mitcham, 2010; Klein, 2010; Krohn, 2010) and it has already provided fruitful concepts and a theoretical background for the analysis of information science's relationships with other disciplines (for a study exploring the interdisciplinary relationships between information science and linguistics see Engerer, 2017b). As this paper is concerned with information science in a supportive, rather than constitutive, role in collaborations the interdisciplinarity perspective will not be discussed further.

Perspective 2 is connected with the activities in a domain (b) and relates information science knowledge to that domain's digitally mediated activities in a project context. From perspective 2, digital communication tools are typically conceptualised as information systems (Boell & Cecez-Kecmanovic, 2015; Urquhart, 2018), distinguished from other system components by their high degree of functional autonomy, input/output features, and knowledge-organising properties (Hjørland, 2003, 2008, 2013) and interactivity (Borlund, 2013; Kiousis, 2002; Ruthven & Kelly, 2011), as well as by their typically dichotomous functional structure. In this structure 'content' is linked to (or 'mediated by') an interpreting receiver, either a human user or a machine (Engerer, 2017a). Under perspective 2 information systems are often regarded as analytically closed systems which can be studied in isolation.

In the following section this line of thought is elaborated and illustrated by an analysis of the Prior website's information systems.

Under perspective 2 the concept of self-contained information systems relates to both intellectual and coordinating labour. Expertise for coordinating functions includes informing the development of project platforms, customising of wikis for information-sharing, integrating conference programmes into a researcher's work desk, embedding tools for announcing project activities into a project website, and developing and maintaining other information systems that directly or indirectly facilitate coordination of a project.

An intellectual labour orientation implies a focus on the knowledge aspect of digital information systems, which serve as media for domain researchers in their cognitive and joint knowledge-building. Information science has a much richer tradition of studying systems for intellectual labour rather than coordinating labour and emphasises that distributed and equal access to domain-relevant information resources and joint terminology are important to the sharing of domain knowledge. One of the characteristics of these systems is a special type of dichotomous functional structure, a nexus linking indexing/metadata and user queries. In the context of intellectual labour information systems deal with both the metadata and representational properties of items of information (Chowdhury, 2010, p. 1; Frohmann, 1990; Lancaster, 2003; Mai, 1999; Svenonius, 2000), typically documents and their retrieval (Baeza-Yates & Ribeiro-Neto, 2011; Pandey, 2003; Ruthven & Kelly, 2011; Warner, 2010). Examples of this kind of dichotomous information system are digital full text repositories, bibliographies, library catalogues, directories of domain-relevant sources, and other digital aids which help researchers to access the information they need directly.

Perspective 3 relates information science knowledge to another kind of knowledge, i.e. practical domain community knowledge (c) about the domain activities (both intellectual and coordinating) of logic and philosophy specialists (b). Practical knowledge of the type c) specifies how things are done in the domain, why they are done this way, and which digital tools are typically used (Wenger, 1998). The first step in connecting information science knowledge with a domain community's practical knowledge is to elicit and externalise tacit domain knowledge, and the second is to map it onto complexes of related information systems in the domain. Hence the question which information science knowledge can be drawn upon has a less straightforward answer than under perspective 2. It is not simply a matter of information systems typology. Faced with the task of eliciting tacit practical knowledge (Nonaka, Toyama, & Konno, 2000), information science can fall back on methods such as domain analysis (Hjørland, 2002), ontology building, user studies, and other qualitative, ethnomethodological techniques (Boaduo, 2011; Daniel, 2011; Pickard, 2013), which can be used to help establish the practical understanding that domain researchers have of their own and their peers' activities. From these approaches we will in the following paragraph elaborate in more detail on domain analysis and ontologies.

As highlighted above, a perspective 3 approach to designing, evaluating, and improving scholarly digital resources builds basically on a correspondence of research community features on the one hand and the structure and design of the information systems used in that community on the other. The more general postulate of a close relationship between user features (e.g., linguistic, social, situational, professional) and the structural makeup of the knowledge systems for these users is in information scientific terms described by the notion of the "domain" (here used in a somewhat narrower sense than in the remainder of this article). A domain from an information science view captures the dependency between knowledge of a specific subject field (such as time logic) and skills in managing and organizing the information resources specifically of that field (such as the Prior Internet Resources [PIR]). Domain analysis takes the view that managing information resources and information systems (databases and websites) in a specific field demands knowledge of this field including its traditions, terminologies, norms, and practices (Albrechtsen, 2015; Bawden & Robinson, 2012; Hjørland, 2002, 2017; López-Huertas, 2015; Robinson, 2009; Tennis, 2003). Birger Hjørland devised a rude but useful and pragmatic methodology (better: systematics) for producing domain-specific knowledge necessary for information scientists in solving their tasks. This methodology includes approaches such as subject gateways, specialist classifications and thesauri, disciplinary peculiars of indexing and retrieving practices, user studies, bibliometrical studies, document and genre studies, terminological studies, historical studies, and more (Hjørland, 2002). Domain-analysis is a practical approach and thus appears as a good starting point for information specialists to systematically collect knowledge about the practices, modes of information seeking, language and communication conventions, and the more in the domains of other disciplines.

A further refinement from a perspective 3 standpoint is ontologies. The information scientific concept of an ontology encompasses the sphere of indexing terms and related search terminology at the same time, and therefore regards index terms as closely linked to (if not identical with) the vocabulary used by specialists in their domain. The step from traditional thesauri

and classification schemes to ontologies of knowledge domains demarcates not only the integration of semantic web principles into the description of data (Berners-Lee, Hendler, & Lassila, 2001), but is foremost a move from barely developing a search terminology towards a controlled language for knowledge representation (Engerer et al., 2017). This step from the lexical-terminological component to a whole language with a built-in logic, a syntax, and inference rules makes it possible to derive information which is not explicitly contained in the descriptive terms themselves (Antoniou, Groth, van Harmelen, & Hoekstra, 2012, p. 4). Knowledge is therefore no longer just named, as it is the case in traditional controlled vocabularies, but it can be described and "confirmed" via ontology languages through constructing sentence-like complex formulas operating with linguistically-informed components such as subjects, verbs (relations), and objects.

Ontologies reflect a common understanding of a domain (Antoniou et al., 2012, p. 11) by expanding the restricted repertoire of thesaural relations between terms (e.g., broader/narrower terms, related terms) to an unrestricted range of semantic relationships realized and acknowledged in the domain idiom. Ontologies must therefore be constructed for each specific domain in order to reflect the language use practiced in the domain in question (Stuart, 2015). Generally, an ontology models the expert user's view on information in his/her domain. More practically, in the case of PIR, the information scientist collects terms, their definitions, and mutual semantic relationships and builds a formal vocabulary system, including syntactical and inference rules. The advantages of ontologies for specialist users are, among others, improved possibilities for exploring data, "semantic search" (King & Reinold, 2008, p. 22), enhanced serendipity, and optimized search results by using ontology-based search techniques including Natural Language Processing (King & Reinold, 2008, p. 12).

Not unlike domain analysis, a complete methodology is linked to the creation of ontologies in specific domains, including steps such as collecting the vocabulary, defining and classifying the vocabulary terms, and indicating the semantic relationships between the established classes (King & Reinold, 2008, ch. 3; Stuart, 2015). Building an ontology is therefore in a way similar to doing a domain analysis as described in Hjørland (2002); both methodologies aim at transferring expert knowledge, often in tacit form, into the realm of explicit knowledge organization, and both respect the linguistic form of this knowledge when modeling it in a knowledge system. Domain analysis and the development of expert ontologies can therefore be regarded as two sides of the same coin, though ontology building is a more specialized activity and part of the broader theoretical endeavors and coverage of domain analysis. The three perspectives are mapped to their respective domain categories a) to c) in Fig. 2.

The illustration highlights two opposite, yet complementary and necessary strands of the process of bringing information science knowledge together with research in other domains (marked with grey). Setting the interdisciplinary perspective (1) aside, we can see that perspective 2 takes an analytical, isolationist approach to information systems, viewing them as self-contained units that are studied by information science in a rather straightforward way, in particular with respect to information systems for intellectual labour. In contrast perspective 3 takes a synthesising, holistic, and integrative approach to information systems, prompted by the tacit, procedural nature of domain communities' knowledge, which

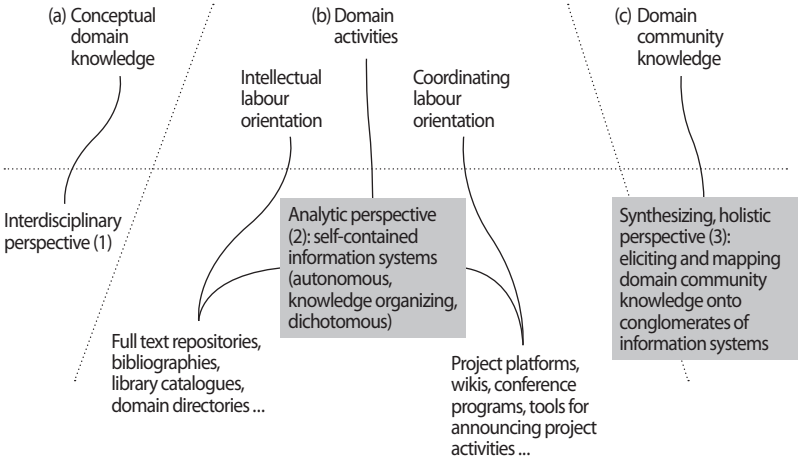


Fig. 2. Three information science perspectives assigned to three knowledge/activity modes of domain researchers.

means that existing practices and workflows must be reflected in the interplay between different information systems (Nicolini, 2013).

Below these two perspectives on information systems, the analytic-isolationist and the synthesising-holistic, are elaborated. It is demonstrated that information scientists need to use both approaches in order to access relevant professional knowledge and respond to domain researchers' two essential ways of constructing their disciplinary domain, namely in terms of what they do (b) and what they know ([a] and [c]).

3. AN ANALYTICAL APPROACH TO INFORMATION SYSTEMS

How is information science knowledge accessed when working with digital research resources? To address this question it is necessary to identify the proper objects for information scientific analysis. The task is, in general terms, to move from functionally unspecified, barely formal organisation units (Internet domains, websites, etc.) (Borgman, 2007) to functionally specialised information systems, which are the appropriate objects of study from an information science point of view (Urquhart, 2018). Once identified these information systems are categorised by type and this typing then serves as a pointer to the disciplinary knowledge that information science has accrued on each. The knowledge associated with each information system type guides analysis of the information system in question and in consequence interventions, improvements in functionality, and other development initiatives can be undertaken in a controlled manner, based on information science analysis.

As one reviewer rightly noted, the dynamic, spatial jargon used here ('move from X to Y') and the arrows in Fig. 3 are based on metaphors with another, grounding concept behind it. In fact, there are four. Firstly the 'moving' metaphor can target the ontology of the information specialist's professional 'world' where he/she starts with given websites, puts focus on the information components in it, names the relevant disciplinary knowledge areas, and accesses them. The 'moving' metaphor is here grounded in a certain chronological concept of professional workflow. Secondly, moving from left to right in the topology represented in Fig. 3 can mean adding accumulatively specificity to digital objects (from websites in general to partial systems with information functions), and cognitive objects (from generic representations of our knowledge of these systems to the knowledge itself). Thirdly, the path from left to right, and, thus, the movement along this path, can be connected and directed by semantic relations represented as arrows in Fig. 3. The arrow from website to information systems connects the two domains by the is-part-of /include relation; knowledge type and knowledge concepts are related by a semiotic reference function, where the type represents, or is a pointer for an information science concept. The transgression from information system to type is more intricate (see the arrow in the middle), as it demarcates the line between two very different spheres, the digital and the cognitive. The assignment of a cognitive knowledge type to a given digital information system is obviously a matter of experience and expertise of practising information specialists; in this sense the assignment of a knowledge representing term to a given information system by an information specialist presupposes an analysis of the latter. Fourthly the outer left and right components in Fig. 3, formal organization and conceptual domain knowledge, can be interpreted as the two sides of the semiotic sign, form and meaning. Accordingly, the movement from formal organization to conceptual knowledge can be interpreted as the systematic assignment of cognitive meanings to formal digital structure. To sum up, moving from left to right can then mean following a workflow, becoming more specific, navigating along semantic relations, or getting closer to the meaning of a digital object. The 'moving' metaphor itself, as used here, is ambiguous with respect to these options.

The first part of this section discusses the theoretical aspects of moving from websites to information systems, and thence to subtypes and the corresponding information science knowledge. The second part illustrates these moves with examples from the PIR. The third part analyses the information systems embedded in the PIR using the theoretical framework set out here.

3.1. From Websites to Domain Knowledge

The path from websites to knowledge via information systems and knowledge types is illustrated in Fig. 3. Below the individual steps are justified and discussed in greater detail.

Inspecting the illustration from left to right we can see that the relationships between formal websites and information systems are not necessarily one-to-one; an information system (in this example information system 3) can formally be distributed over two or more web domains (indicated by a circle overlapping with two websites) and, perhaps trivially, one website can contain more than one information system, as it is the case with website 1, in which information systems 1 and 2 are embedded. The relationships between websites and information systems connect the digital, formal organisational level with the information systems level, marking the transition from a rather non-specific domain of communication to a functionally more

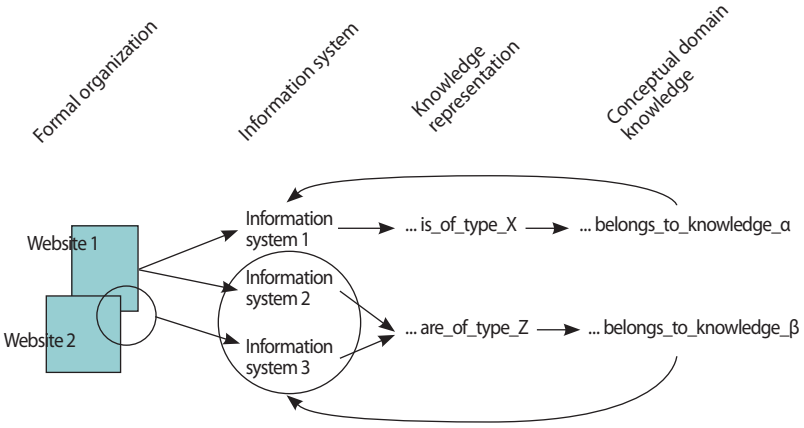


Fig. 3. From websites to information systems, to types of information systems, to type-specific knowledge, and back to the information system.

specific domain, in our case the informational domain.

Looking further to the right of the illustration we reach the knowledge representation level (Blair, 1990; Stock & Stock, 2013), where the type of information system is specified. Again, a non-unique relationship connects knowledge types with information systems. One information system can be related to just one type of knowledge (in this example information system 1 is connected only with knowledge type X), or two information systems (here, 2 and 3) may be instantiations of the same knowledge type (here, type Z). Note that under perspective 2 the knowledge types (and hence also the domain knowledge) are assigned to the information systems regardless of 1) how many instantiations they possess on the level of information systems and 2) their formal organisational properties, which is the level at which the context of these systems is determined (in other words, the context in which an information system is embedded is irrelevant to the assignment of knowledge types). Clearly, therefore, perspective 2 is unsatisfactory, and perspective 3, which is complementary, compensates for this, as shown below.

Moving from information systems to knowledge representations is a step towards domain specialisation. This specialisation is achieved by mapping non-specific information systems onto an information scientific nomenclature for domain terms. This acknowledgement of the disciplinary terminology of a knowledge-organising system (Temmerman, 2000), not unlike a thesaurus system (Broughton, 2006; Foskett, 1994; Lykke Nielsen, 2001), is the key to the conceptual domain-specific knowledge.

The right side of Fig. 3 shows the linking of conceptual domain-specific knowledge (more commonly termed 'expert knowledge' or 'expertise') and information system types. Relationships between knowledge-type terms and the knowledge system itself are also not unambiguous (although this is not indicated in the illustration). In practice an information system type can map to more than one knowledge system or concept; similarly, two or more different information system types can relate to the same kind of knowledge system or concept. Theories of knowledge organisation treat these relationships as ambiguous descriptor terms in a thesaurus system (our knowledge types) with several, often incompatible 'meanings,' 'definitions,' 'referents,' 'semantics,' etc. (our conceptual domain-specific knowledge). In traditional information organisation, ambiguities like the ones mentioned above are treated as mismatches and undesirable drawbacks (Svenonius, 2000). In our very different context we might simply conclude that the information professional faces the challenge of identifying the most appropriate domain knowledge for the information system type under inquiry.

Once the relevant information science knowledge is identified, it must be projected into the realm of information systems where it guides analysis and, eventually, prompts interventions. In our illustration, the transfer of partial information science knowledge into information systems is indicated by backwards arrows.

3.2. Illustration of Moves Using the PIR

In this section the abstract moves presented in the previous section are illustrated using cases taken from the PIR. The numbers 1-6 designate the six information systems in the PIR, as illustrated in Fig. 4. Methodologically, the identification of information systems (3.2.1) differs from the identification of knowledge types (3.2.2) and conceptual knowledge (3.2.3). While the status of an information system is mainly a matter of design affordances, web site architecture and, most importantly, the way user practices apply to these affordances, the identification of knowledge types and concepts in relation to these systems is guided by disciplinary, information science meta knowledge regarding how to assign a knowledge type to an existing information system and connect this type with the relevant knowledge domains of the discipline. Thus, on the one hand, whether a website facility functions as an information system or not is to a high degree dependent on the user practices, here the uses that time logicians (logicians working on time as a philosophical subject), philosophers, historians, and more make of a website. On the other hand, assigning information science knowledge types to digital systems and accessing the corresponding information science concepts is solely based on information science practice and method.

We can therefore methodologically ground the move from websites to information systems (3.2.1) by addressing design features, website architecture characteristics, and the information practices of the website's user group. Decisions concerning types and concepts (3.2.2 and 3.2.3), however, must methodologically be based on information science practice, i.e. the ways how information specialists draw on their professional knowledge in developing information systems. With regard to the former, knowledge from interaction design (for example, Preece, Sharp, & Rogers, 2015) and information architecture (for instance, Ding, Lin, Zarro, & Marchionini, 2017) is coupled with informal assessments of the website and its information systems by members of the domain group (collected in informal conversations). With regard to the latter case, in which information professionals access the relevant disciplinary knowledge they need to apply to a digital system, only introspective reflection into their own professional practices reveals their competencies and practical skills. These reflections have been discussed in the information science group of the project and have been discussed by the information science group in several publications (Engerer et al., 2017; Engerer & Albretsen, 2017; Engerer & Sabir, 2018). What we do here, in a nutshell, is to make explicit *a posteriori* our own tacit model of accessing and applying information science disciplinary knowledge in our concrete practice of website development. In this sense, other researchers may arrive at other 'models' in their own practice, but it is finally this awareness and explicitness that is the basis for development and improvement in professional environments.

3.2.1. From Websites to Information Systems

As stated above, one information system can be distributed over two or more web domains and vice versa, i.e. one website

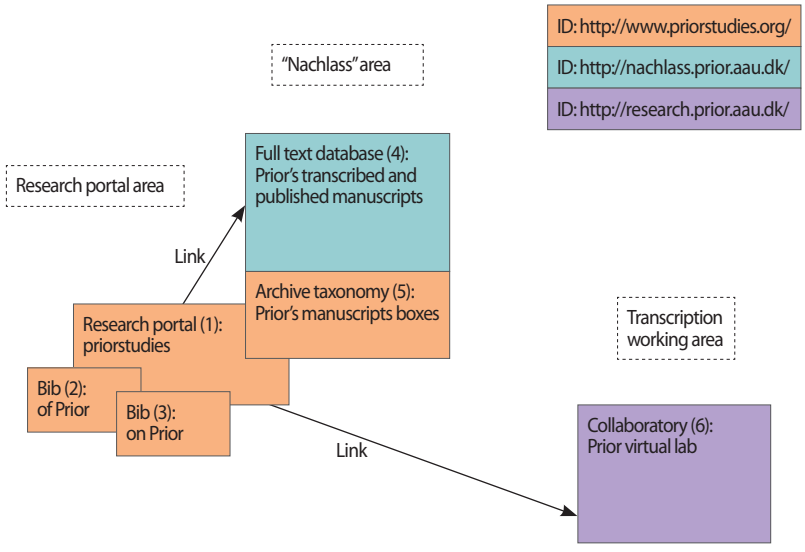


Fig. 4. General structure of Prior Internet Resources (as of summer 2017): six information systems representing five distinct information system types, implemented on three Internet domains functioning in three areas.

can contain more than one information system. An example of the latter situation is the Prior Studies website,[3] which embeds two bibliographies (Of Prior, On Prior, no. 2 and 3) and also contains the archive taxonomy as a functional part of the Nachlass area (no. 5). An interesting illustration of the former situation is the complex Nachlass area (depicted in Fig. 4), which covers handwritten material by Prior from the Bodleian Library archives and is, as seen from perspective 3 and through the lens of the research community's practices, one information system unit. From perspective 2 however, it encompasses three information systems: the archive taxonomy (no. 5, a classificatory entrance to the material), the full text database containing manuscripts which have already been transcribed and published (no. 4), and the Prior Virtual Lab (no. 6), which contains photographic images of unpublished handwritten material. All three systems are distributed over three different Internet domains and embedded in completely different websites and contexts. This makes it difficult for users to recognise the Nachlass area's functional coherence.

There is an obvious discrepancy between larger, functional units such as the Nachlass area and the rather isolationist construct of the information systems that constitute them. The latter focuses on single information systems and their input and output characteristics, whereas the former highlights the interplay between the systems and how, collectively, they serve the research practices of the users.

3.2.2. From Information Systems to Knowledge Types

The relationship between information systems and knowledge types is also variable. One information system can be related to one knowledge type and two information systems can instantiate a single knowledge type. An example of the former, straightforward case is provided by the full-text repository of transcribed and published Prior manuscripts (no. 4), which can be classed unequivocally as a full-text database. One example of the latter case would be two information systems Of Prior and On Prior (no. 2 and 3), which are lists of Prior's writings and secondary literature which has Prior as its subject. Clearly, in the light of our subjective, introspective methodology, one can express doubt about the status of the two bibliographies. Could users, as one reviewer asked, not conceptualize these as one composite information system or maybe not even consider them separate at all? In our conversations with Prior scholars, however, we realized that access to the original writings by Arthur Prior was very differently conceptualized and had very different status for the project participants than access to secondary literature about Prior. This has, among other things, to do with a speciality in the project research practices, namely regarding Prior's writings as, in a sense, primary objects of research and interpretation *per se*, while research and interpretation is documented in 'secondary' publications. We also want to mention the pivotal role of Arthur Prior as a person in general for the project. Thus, in spite of these user-related, practice-based distinctions between Of Prior and On Prior, these two information systems are of the same information science knowledge type, namely bibliography.

3.2.3. From Knowledge Types to Conceptual Knowledge

As already noted, one information system type can correspond to more than one knowledge system or concept as, for example, in the case of the Prior Virtual Lab (no. 6), where the information system type 'collaboratory' (Bos et al., 2008; Finholt, 2002) points at several complementary theories or approaches, depending on whether the focus is on transcription labour, communication, and knowledge sharing among the participants or the organisation and availability of the photographed manuscripts. On the other hand, two or more different information system types can be related to the same kind of knowledge system or concept. This kind of ambiguity is illustrated by the information system types bibliography (no. 2 and 3), full-text database (no. 4), and archive taxonomy (no. 5), which all draw on concepts such as metadata (Hider, 2012), indexing categories (Lancaster, 2003; Weinberg, 2009), classification (Batley, 2005), and taxonomic principles (Bawden & Robinson, 2012).

3.3. Information Systems in the PIR

The analytic approach will now be illustrated using the PIR, which comprise the Internet resources on Arthur Prior that are associated with the Prior project. The term PIR encompasses both formal digital elements such as websites, knowledge-organising units such as bibliographies, and other information systems, which will be presented in further detail below. The overall structure of the PIR, depicted in Fig. 4, consists of three main components, i.e. 'content areas' (these do not strictly coincide with particular Internet domains). 'Foundations of Temporal Logic—The WWW-site for Prior-studies' (Hasle & Øhrstrøm, 2016), hereafter 'Priorstudies,' is the main entry point (research portal) for scholars interested in Arthur Prior's work and life. The related 'Virtual Lab for Prior Studies'

[3] The Danish Prior Internet representation has been revised several times in the years 2018/19, and the result of this development process can be accessed under http://www.priorstudies.org. Our own analysis refers to the website as it was before these improvements (which are connected with the work done by the author) until 2017. This 'historical' website can be accessed through http://web.archive.org/web/20070609124540/http://www.kommunikation.aau.dk/prior/index2.htm, which is an archived version.

(Albretsen, 2016), hereafter Prior Virtual Lab (PVL), is the virtual platform used by researchers transcribing handwritten documents by Prior. Finally, we have the so-called 'Nachlass,' a full-text archive of transcribed and published Prior manuscripts (Nachlass area). As already mentioned, our analysis refers to the Prior sites as they were until 2017, not taking into account subsequent changes and design modifications. This older version is accessible and archived at http://web.archive.org/web/20070609124540/http://www.kommunikation.aau.dk/prior/index2.htm.

As indicated in the preceding section, the information systems have to be identified first. The term 'information system,' which has its roots in the world of management and business (Burton Swanson, 2009), refers by default to IT-based support to enable organisations to accomplish specific tasks (cf. Wallace, 2015), but definitions and conceptions of information systems vary considerably according to whether the perspective is technological, social, sociotechnical, or process-oriented, as a thorough review of information systems' definitions (Boell & Cecez-Kecmanovic, 2015) demonstrated. Perhaps the broadest definition of an information system—but still meaningful in the context of this paper—is the one that comes from Wikipedia; it is widely cited in textbooks (for example, Bourgeois, 2014) and on conference websites (in the Wikipedia article itself no references are given). According to this definition "[a]n information system […] is an organised system for the collection, organisation, storage and communication of information. More specifically, it is the study of complementary networks that people and organisations use to collect, filter, process, create and distribute data." (Wikipedia, 21 April 2017, https://en.wikipedia.org/wiki/Information_system, link marking and bold type removed).

This broad definition of information systems is practical and allows the approximate identification of six information systems embedded in the PIR.[4] Each information system in the PIR has been glossed in the list below with the specific information-related action associated with it, taken from the definition above:

1) 'Foundations of Temporal Logic—The WWW-site for Prior-studies': e.g., communication of information
2) Works written by Prior, primary literature: e.g., collection, organisation of information
3) Works written on Prior, secondary literature: same as 2)
4) 'Nachlass' (full-text): e.g., organised system for the collection, organisation, storage and communication of information
5) 'Nachlass' in the archive boxes: e.g., organisation of information
6) Prior Virtual Lab: e.g. complementary networks that people and organisations use to collect, filter, process, create and distribute data

In information science, information systems are of several types—most prominently documentary languages implemented in knowledge organisation systems such as classification systems, thesauri, and ontologies (Hjørland, 2003, 2013; Stock & Stock, 2013, sect. L); information services such as bibliographies, retrievable databases, and text repositories; and, last but not least, research portals and collaborative academic platforms in general. In order to identify the information science knowledge relevant to the six PIR-embedded information systems, these systems have to be mapped onto specific types of information systems, such as the ones mentioned. The goal of this exercise is to enable systematic access to relevant and useful scientific disciplinary knowledge, which improves our understanding of the PIR and can be a starting point for professionals seeking to develop and improve the existing digital resources.

The PIR, defined as the virtual space delimited by the three abovementioned content areas and Internet domains (Priorstudies, PVL, and Nachlass), contains, on first inspection,[5] six information systems of five distinct types. All information systems types are well-known and acknowledged in the information science research tradition, and disciplinary knowledge relevant to each type is readily accessible:

- Information system 1: 'Foundations of Temporal Logic—The WWW-site for Prior-studies' (part of the Priorstudies Internet domain); type: research portal; exemplary domain knowledge (Becker et al., 2012; Elsayed et al., 2011)
- Information system 2: 'Of Prior,' works written by Prior, primary literature (part of the Priorstudies Internet

[4] It has to be emphasised that the Wikipedia definition is by no means sufficient as an operational definition, nor does it really explain what information systems are. Clearly, there is much heuristic supposition involved in the identifications given above, but if these six objects can be mapped onto significant information systems types and thus be meaningfully and instructively linked to information science knowledge they should provide a better understanding of these systems and thus provide practical confirmation of the plausibility of the initial decisions.

[5] Again, it has to be emphasised that there is a great deal of heuristic assumption involved in specifying the types of knowledge systems that exist in the information science domain and how they can be recognised in a variety of instantiations of digital information systems. The same can be said about the assignment of information science concepts to information system types. Clearly, professional background, professional experience, and theoretical inclinations play a crucial role in determining the theories and works an information scientist draws upon when he/she describes a specific type of information system.

domain); type: bibliographical database; exemplary domain knowledge (Chowdhury, 2010, p. 17; Hider, 2012)
- Information system 3: 'On Prior,' works on Prior, secondary literature (part of the Priorstudies Internet domain); type and exemplary domain knowledge is the same as in information system 2
- Information system 4: 'Nachlass' in its narrow meaning (Nachlass Internet domain); type: full-text database, text repository; exemplary domain knowledge (Borgman, 2007; Eggert, 2009; Lin, Fan, & Zhang, 2009; Littlejohn, 2005)
- Information system 5: 'Nachlass' in the archive boxes (part of the Priorstudies Internet domain); type: taxonomic entry for archival metadata; exemplary domain knowledge (Batley, 2005; Bawden & Robinson, 2012; Broughton, 2006; Millar, 2017; Thomas, Fowler, & Johnson, 2017)
- Information system 6: Prior Virtual Lab (Prior Virtual Lab Internet domain); type: collaboratory, research platform; exemplary domain knowledge (Becker et al., 2012; Bos et al., 2008; Elsayed et al., 2011; Finholt, 2002)

At the 2017 stage of the Prior project the digital information structure of the PIR consists of four theoretical levels: formal organisation (three Internet domains), information systems (six partial systems), knowledge representation (five information science subtypes), and conceptual domain-specific knowledge (five partial knowledge domains, corresponding to five information science subtypes). At the time of writing a complete restructuring of the website is in progress; some preliminary results are presented in Engerer and Albretsen (2017). Fig. 4 sketches the general structure of PIR in summer 2017.

4. A SYNTHETIC APPROACH TO INFORMATION SYSTEMS: THE 'TRANSCRIBER LOOP'

In this final section the perspective 3 standpoint is explored further (although rather informally and in less detail than for perspective 2). This discussion 'corrects' the flaws of the analytical view of information systems and emphasises the fact that such systems are always part of a larger system which must support existing practices and workflow in the domain (Nicolini, 2013).

From the perspective of domain users there are substantial functional connections between the three information systems making up the PIR as constituted at the time of writing. The dynamics of the relationship between the archive taxonomy (no. 5, Prior's archive boxes), the PVL (no. 6), and the Nachlass full-text database (no. 4, transcribed and published manuscripts by Prior) is crucial to the work of the PVL (a collaboratory, no. 6), i.e., transcribing, digitising, and making accessible via the Internet as many unpublished manuscripts by Prior as possible. The role of the box taxonomy from the Nachlass section is particularly important, as it is the only point of departure for Prior scholars seeking to match topics in the original handwritten material with their own research questions and research interests.

It is important to note that at that point of their inquiry researchers do not have the opportunity to verify documents' relevance by consulting an electronic copy of the original paper in the archive (Blair & Kimbrough, 2002). They only have access to documents' metadata, their representations, which the researcher has to treat as reliable surrogates for the original document. A preliminary match should give a scholar an incentive to sign up to use the PVL, request a copy, and then determine whether the text is worth transcribing. In other words, if Prior scholars cannot reliably ascertain whether the archive boxes contain documents relevant to their research questions, it is highly unlikely that they will register to use the PVL.

The box taxonomy must therefore be viewed as the hub of the transcription project. It is where the researcher initiates a document cycle, the starting point of which is an attempt to identify a document that is suitable for transcription; should the attempt be successful the document is transcribed and eventually returns to the Nachlass as a full-text, searchable electronic document, provided with a fully-fledged set of metadata. This dynamics of the relationship between information systems, under one system umbrella, and domain user properties is illustrated below in Fig. 5, which shows the pathways of researchers and manuscripts/documents between the three information systems schematically, in the form of a loop, the researcher-to-document loop or, more succinctly, the transcriber loop.

In the initial phase of the manuscript cycle, the 'manuscript-born' index fields, which have been derived from a specialist archiver's descriptions, act as a kind of 'beginner set,' attached to the handwritten text. They are extraordinarily valuable access points (Hjørland, 1998; Lancaster, 2003, p. 6) for advanced specialist searches. As the illustration shows, the researcher then assumes the role of a transcriber. In this transition the initial archiver's metadata accompany the manuscript. In this stage the researcher not only carries out the transcription, but also enriches the manuscript metadata from the archive with information drawn from his/her expert knowledge and

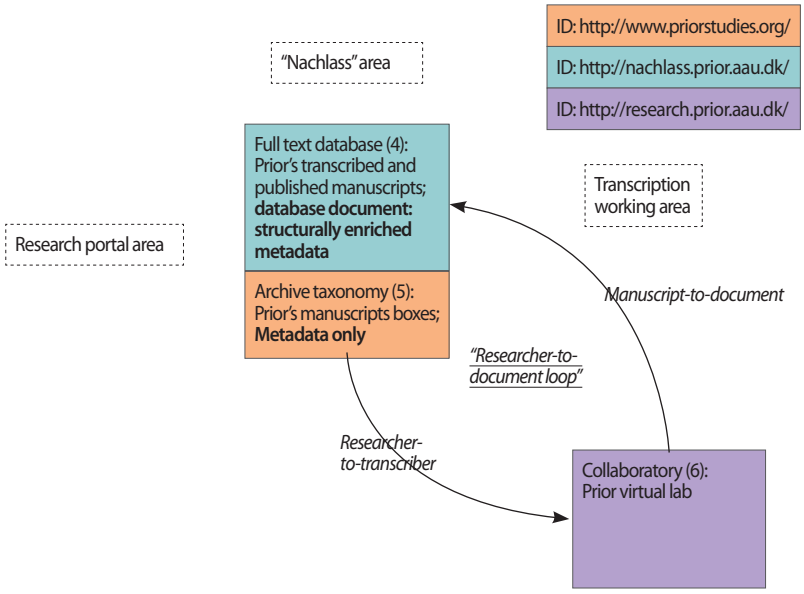


Fig. 5. The researcher-to-document loop connecting three information systems in the Prior Internet Resources.

textual or contextual knowledge arising from his/her deep intellectual engagement with the content of the manuscript during the transcription process. This is an important aspect of the manuscript-to-document process, indicated by the arrow from the PVL to the Nachlass full-text database. The sequential aggregation of metadata, as shown here, is a typical case of 'enrichment via information-added values,' whereby texts are formally described and indexed for content, resulting in fully-fledged surrogates, sometimes called 'documentary units' (Stock & Stock, 2013, p. 69).

The last step of the manuscript-to-document process is the formal adaptation of the documentary units to a database environment, a formally organised collection of surrogates which can be searched, retrieved, and explored. This makes them what often is called a 'record.' From this information science perspective, the manuscript-to-document arrow signifies a text's change of status from a more or less unstructured and informal piece of text to a standardised record in a formal, machine-readable, and searchable database in the full-text Nachlass. Processes such as these can only be understood when an account of the interactions of information systems, linked to profound knowledge of the domain group's practices and motivational factors, supplements the analytical approach to information systems as illustrated in the previous sections.

5. CONCLUSION

This paper explores how information science knowledge can contribute to research collaborations in which information scientists support colleagues from another discipline. Information science comes into contact with domain research in three forms: knowledge of the domain (conceptual knowledge), intellectual and coordinating functions (activities in the domain), and knowledge of how things are done in the domain and why (practical knowledge). These three forms engender three perspectives on information science knowledge: 1) the interdisciplinary view (not covered in this paper), 2) an analytical and isolationist perspective on information systems, and 3) a synthetic and holistic perspective, which sees information systems as interacting units responding to practices in the domain. From perspective 2, information science knowledge is accessed in three moves: from websites to information systems, from information systems to information system types, and finally from information system types to knowledge systems which can be used to develop the information systems in question. These moves have been illustrated through an analysis of the PIR. The synthetic perspective is exemplified by the transcriber loop, in which researchers move across three information systems, the box taxonomy, the PVL, and the full-text Nachlass, in order to

execute a complex task, namely transcribing and commenting on a photographic image from the collection of handwritten manuscripts by Prior.

The paper makes the point that, in order to access the knowledge they need when working with information systems, it is important for information professionals to understand whether they react to activities in the domain (perspective 2) or to tacit knowledge of these activities, for example the structure and motivation of workflows (perspective 3). In the first case the path to professional knowledge is straightforward and flows from information system to information system type to information science knowledge. In the latter case an intermediate step, eliciting domain researchers' practical knowledge of their workflows, is required; this knowledge can then be mapped onto a more complex structure of interdependent information systems.

There are, therefore, two intended audiences for this paper. The first consists broadly of information professionals, i.e. research librarians, information specialists working with information systems in domain-specific settings, and computer scientists. The paper offers this audience concrete help in identifying the appropriate strategy for handling professional knowledge in order to take appropriate actions and development decisions. Defining one's own position towards researchers' doings and knowings in the domain makes it easier to determine what support one should offer. It helps to clarify an originally vague and ambiguous situation.

The paper's other target audience is senior researchers writing research proposals and principle investigators already engaged in research project management, including decisions about resource allocation. The paper provides this audience with criteria for describing precisely the nature of a digital project's resources, according to whether the priority is development of project-specific information and communication systems (perspective 2) or the mapping of larger tasks, such as manuscript transcription, onto conglomerates of digital systems (perspective 3). Knowledge of the distinctions presented here will support deliberate and selective allocation of project resources and could inform the decision about which types of information professional should be recruited to assist with the project. Such knowledge should help to define shared expectations and make collaboration smoother and more effective.

ACKNOWLEDGMENTS

This paper is based on research in the project "The Primacy of Tense: A. N. Prior Now and Then," funded 2016-2019 by the Danish Council for Independent Research-Humanities (DFF-FKK Grant-ID: DFF 6107-00087).

REFERENCES

Albrechtsen, H. (2015). This is not domain analysis. *Knowledge Organization*, 42(8), 557-561.

Albretsen, J. (2016). *The virtual lab for prior studies*. Retrieved April 10, 2019 from http://research.prior.aau.dk/login_user.php.

Antoniou, G., Groth, P., van Harmelen, F., & Hoekstra, R. (2012). *A semantic web primer*. Cambridge: MIT Press.

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern information retrieval: The concepts and technology behind search*. New York: Addison Wesley.

Batley, S. (2005). *Classification in theory and practice*. Oxford: Chandos Publishing.

Bawden, D., & Robinson, L. (2012). *An introduction to information science*. London: Facet.

Becker, J., Knackstedt, R., Lis, L., Stein, A., & Steinhorst, M. (2012). Research portals: Status quo and improvement perspectives. *International Journal of Knowledge Management*, 8(3), 27-46.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 28-37.

Blair, D. C. (1990). *Language and representation in information retrieval*. Amsterdam: Elsevier Science.

Blair, D. C., & Kimbrough, S. O. (2002). Exemplary documents: A foundation for information retrieval design. *Information Processing & Management*, 38(3), 363-379.

Boaduo, N. A. (2011). Research methods for studying virtual communities. In B. K. Daniel (Ed.), *Handbook of research on methods and techniques for studying virtual communities: Paradigms and phenomena* (pp. 549-610). Hershey: Information Science Reference.

Boell, S. K., & Cecez-Kecmanovic, D. (2015). What is an information system? In *2015 48th Hawaii International Conference on System Sciences*, Kauai, HI, USA (pp. 4959-4968). Piscataway: IEEE.

Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the internet*. Cambridge: MIT Press.

Borlund, P. (2013). Interactive information retrieval: An introduction. *Journal of Information Science Theory and Practice*, 1(3), 12-32.

Bos, N., Zimmerman, A., Olson, J. S., Yew, J., Yerkie, J., Dahl, E., . . . Olson, G. M. (2008). From shared databases to

communities of practice: A taxonomy of collaboratories. In G. M. Olson, A. Zimmerman, & N. Bos (Eds.), *Scientific collaboration on the Internet* (pp. 53-72). Cambridge: MIT Press.

Bourgeois, D. T. (2014). *Information systems for business and beyond*. Retrieved April 10, 2019 from https://bus206.pressbooks.com/front-matter/introduction.

Broughton, V. (2006). *Essential thesaurus construction*. London: Facet.

Burton Swanson, E. (2009). Information systems. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences* (pp. 2635-2642). Boca Raton: CRC Press.

Case, D. O. (2012). *Looking for information: A survey of research on information seeking, needs and behavior*. Bingley: Emerald.

Chowdhury, G. G. (2010). *Introduction to modern information retrieval*. London: Facet.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning*, 3(2), 105-122.

Cummings, J. N., & Kiesler, S. (2008). Collaborative research across disciplinary and organizational boundaries. In G. M. Olson, A. Zimmerman, & N. Bos (Eds.), *Scientific collaboration on the Internet* (pp. 99-117). Cambridge: MIT Press.

Daniel, B. K. (2011). *Handbook of research on methods and techniques for studying virtual communities: Paradigms and phenomena*. Hershey: Information Science Reference.

Dervin, B., Foreman-Wernet L., & Lauterbach, E. (2003). *Sense-making methodology reader, selected writings of Brenda Dervin*. Cresskill: Hampton Press.

Ding, W., Lin, X., Zarro, M., & Marchionini, G. (2017). *Information architecture, the design and integration of information spaces*. San Rafael: Morgan & Claypool Publishers.

Eggert, P. (2009). The book, the E-text and the 'work-site'. In M. Deegan & K. Sutherland (Eds.), *Text editing, print and the digital world* (pp. 63-82). Farnham: Ashgate.

Elsayed, I., Madey, G., & Brezany, P. (2011). Portals for collaborative research communities: Two distinguished case studies. *Concurrency and Computation: Practice and Experience*, 23(3), 269-278.

Engerer, V. (2017a). Control and syntagmatization: Vocabulary requirements in information retrieval thesauri and natural language lexicons. *Journal of the Association for Information Science and Technology*, 68(6), 1480-1490.

Engerer, V. (2017b). Exploring interdisciplinary relationships between linguistics and information retrieval from the 1960s to today. *Journal of the Association for Information Science and Technology*, 68(3), 660-680.

Engerer, V., & Albretsen, J. (2017). The Prior internet resources 2017: Information systems and development perspectives. In P. F. V. Hasle, P. Blackburn, & P. Øhrstrøm (Eds.), *Logic and philosophy of time: Themes from Prior* (pp. 223-249). Aalborg: Aalborg University Press.

Engerer, V., & Gudiksen, J. (2016). *From cognition to practice: theoretical perspectives on the relationship between disciplinary learning and information seeking*. Unpublished manuscript. Retrieved April 10, 2019 from https://www.academia.edu/11969257/From_cognition_to_practice_theoretical_perspectives_on_the_relationship_between_disciplinary_learning_and_information_seeking.

Engerer, V., Roued-Cunliffe, H., Albretsen, J., & Hasle, P. F. V. (2017). The prior-project: From archive boxes to a research community. In *Digital Humanities in the Nordic Countries, Gothenburg*, Sweden, March 14-16 (pp. 53-57).

Engerer, V., & Sabir, F. (2018). Information professionals meet Arthur Prior. *Journal of Librarianship and Information Science*. doi:10.1177/0961000618799527.

Erkens, G., Prangsma, M. E., & Jaspers, J. G. M. (2006). Planning and coordinating activities in collaborative learning. In A. M. O'Donnell, C. E. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology* (pp. 233-263). Mahwah: L. Erlbaum Associates.

Finholt, T. A. (2002). Collaboratories. *Annual Review of Information Science and Technology*, 36(1), 73-107.

Flanders, J. (2012). Collaboration and dissent: Challenges of collaborative standards for digital humanities. In M. Deegan & W. McCarty (Eds.), *Collaborative research in the digital humanities* (pp. 67-80). Farnham: Ashgate Pub.

Floridi, L. (2011). *The philosophy of information*. Oxford: Oxford University Press.

Foskett, D. J. (1994). Thesaurus. In K. Sparck Jones & P. Willet (Eds.), *Readings in information retrieval* (pp. 111-134). San Francisco: Morgan Kaufmann.

Frodeman, R., Klein, J. T., & Mitcham, C. (2010). *The Oxford handbook of interdisciplinarity*. Oxford: Oxford University Press.

Frohmann, B. (1990). Rules of indexing: A critique of mentalism in information retrieval theory. *Journal of Documentation*, 46(2), 81-101.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, 83(1), 88-96.

Gurak, L. J., & Antonijevic, S. (2012). The psychology of blogging: You, me, and everyone in between. In J. Hughes (Ed.), *SAGE internet research methods* (pp. v4-93-v4-103). London: Sage.

Hasle, P. F. V., & Øhrstrøm, P. (2016). *Foundations of temporal logic: The WWW-site for Prior-studies*. Retrieved April 10, 2019 from http://www.priorstudies.org.

Hider, P. (2012). *Information resource description: Creating and managing metadata*. London: Facet.

Hjørland, B. (1998). Information retrieval, text composition, and semantics. *Knowledge Organization*, 25(1-2), 16-31.

Hjørland, B. (2002). Domain analysis in information science: Eleven approaches: Traditional as well as innovative. *Journal of Documentation*, 58(4), 422-462.

Hjørland, B. (2003). Fundamentals of knowledge organization. *Knowledge Organization*, 30(2), 87-111.

Hjørland, B. (2008). What is knowledge organization (KO)? *Knowledge Organization*, 35(2-3), 86-101.

Hjørland, B. (2013). Theories of knowledge organization: Theories of knowledge. *Knowledge Organization*, 40(3), 169-181.

Hjørland, B. (2017). Domain analysis. *Knowledge Organization*, 44(6), 436-464.

Hockey, S. (2012). Digital humanities in the age of the internet: Reaching out to other communities. In M. Deegan & W. McCarty (Eds.), *Collaborative research in the digital humanities* (pp. 81-92). Farnham: Ashgate Pub.

Kimmerle, J., Cress, U., & Moskaliuk, J. (2012). Wiki-supported knowledge building - theory, research, and application. In S. Bocconi & G. Trentin (Eds.), *Wiki supporting formal and informal learning* (pp. 41-55). Hauppauge: Nova Science Publishers.

King, B. E., & Reinold, K. (2008). *Finding the concept, not just the word, a librarian's guide to ontologies and semantics*. Oxford: Chandos.

Kiousis, S. (2002). Interactivity: A concept explication. *New Media & Society*, 4(3), 355-383.

Klein, J. T. (2010). A taxonomy of interdisciplinarity. In R. Frodeman, J. T. Klein, & C. Mitcham (Eds.), *The Oxford handbook of interdisciplinarity* (pp. 15-30). Oxford: Oxford University Press.

Krohn, W. (2010). Interdisciplinary cases and disciplinary knowledge. In R. Frodeman, J. T. Klein, & C. Mitcham (Eds.), *The Oxford handbook of interdisciplinarity* (pp. 31-49). Oxford: Oxford University Press.

Lancaster, F. W. (2003). *Indexing and abstracting in theory and practice*. London: Facet.

Lin, H., Fan, W., & Zhang, Z. (2009). A qualitative study of web-based knowledge communities: Examining success factors. *International Journal of E-Collaboration*, 5(3), 39-57.

Littlejohn, A. (2005). Key issues in the design and delivery of technology-enhanced learning. In P. Levy & S. Roberts (Eds.), *Developing the new learning environment: The changing role of the academic librarian* (pp. 70-90). London: Facet.

López-Huertas, M. J. (2015). Domain analysis for interdisciplinary knowledge domains. *Knowledge Organization*, 42(8), 570-580.

Lykke Nielsen, M. (2001). A framework for work task based thesaurus design. *Journal of Documentation*, 57(6), 774-797.

Mai, J. (1999). Deconstructing the indexing process. *Advances in Librarianship*, 23, 269-298.

Millar, L. A. (2017). *Archives: Principles and practices*. London: Facet.

Nicolini, D. (2013). *Practice theory, work, and organization, an introduction*. Oxford: Oxford University Press.

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, ba and leadership: A unified model of dynamic knowledge creation. *Long Range Planning*, 33(1), 5-34.

Notari, M., & Honegger, B. D. (2012). Wiki: An archetypical tool for collaborative learning in the digital age. In S. Bocconi & G. Trentin (Eds.), *Wiki supporting formal and informal learning* (pp. 21-39). Hauppauge: Nova Science Publishers.

Olson, J. S., Hofer, E. C., Bos, N., Zimmerman, A., Olson, G. M., Cooney, D., & Faniel, I. (2008). A theory of remote scientific collaboration. In G. M. Olson, A. Zimmerman, & N. Bos (Eds.), *Scientific collaboration on the Internet* (pp. 73-97). Cambridge: MIT Press.

Østerlund, C., Snyder, J., Sawyer, S., Sharma, S., & Willis, M. (2015). Documenting work: From participant observation to participant tracing. In K. D. Elsbach & R. M. Kramer (Eds.), *Handbook of qualitative organizational research* (pp. 391-400). New York: Routledge.

Pandey, R. C. (2003). *Information retrieval system: A linguistic study*. Delhi: Abhijeet Publications.

Pickard, A. J. (2013). *Research methods in information*. London: Facet.

Preece, J., Sharp, H., & Rogers, Y. (2015). *Interaction design, beyond human-computer interaction*. Chichester: Wiley.

Prior Project Group. (2017). *The Primacy of Tense: A. N. Prior Now and Then*. Retrieved April 10, 2019 from http://www.prior.aau.dk.

Ramage, M. A. (2010). Evaluating collaborative technologies: A simple method. In H. M. Donelan, K. L. Kear, & M. A. Ramage (Eds.), *Online communication and collaboration:*

A reader (pp. 73-77). New York: Routledge.

Robinson, L. (2009). Information science: Communication chain and domain analysis. *Journal of Documentation*, 65(4), 578-591.

Robinson, L., & Schulz, J. (2012). New avenues for sociological inquiry: Evolving forms of ethnographic practice. In J. Hughes (Ed.), *SAGE internet research methods* (pp. v4-33-v4-47). London: Sage.

Rolland, B., & Potter, J. D. (2017). On the facilitation of collaborative research: Enter stage left, the consortium director. *Cancer Epidemiology, Biomarkers & Prevention*, 26(11), 1581-1582.

Ruthven, I., & Kelly, D. (2011). *Interactive information seeking, behaviour and retrieval*. London: Facet.

Schatzki, T. R. (1996). Social practices: *A Wittgensteinian approach to human activity and the social*. Cambridge: Cambridge University Press.

Stock, W. G., & Stock, M. (2013). *Handbook of information science*. Berlin: deGruyter Saur.

Stuart, D. (2015). *Practical ontologies for information professionals*. London: Facet.

Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge: MIT Press.

Temmerman, R. (2000). *Towards new ways of terminology description: The sociocognitive-approach*. Amsterdam: J. Benjamins.

Tennis, J. T. (2003). Two axes of domains for domain analysis. *Knowledge Organization*, 30(3), 191-195.

Thomas, D., Fowler, S., & Johnson, V. (2017). *The silence of the archive*. London: Facet.

Tompkins, P., Perry, S., & Lippincott, J. K. (1998). New learning communities: Collaboration, networking, and information literacy. *Information Technology and Libraries*, 17(2), 100-106.

Tredinnick, L. (2007). Digital information culture: The individual and society in the digital age. Oxford: Chandos.

Urquhart, C. (2018). *Information systems, process and practice*. London: Facet.

Vasileiadou, E. (2012). Research teams as complex systems: Implications for knowledge management. *Knowledge Management Research & Practice*, 10(2), 118-127.

Wallace, P. M. (2001). *The psychology of the Internet*. Cambridge: Cambridge University Press.

Wallace, P. M. (2015). *Introduction to information systems*. Boston: Pearson.

Warner, J. (2010). *Human information retrieval*. Cambridge: MIT Press.

Weinberg, B. H. (2009). Indexing: History and theory. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences* (pp. 2277-2290). New York: Marcel Dekker.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.

Zaugg, H., West, R. E., Tateishi, I., & Randall, D. L. (2011). Mendeley: Creating communities of scholarly inquiry through research collaboration. *TechTrends*, 55(1), 32-36.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice
eISSN : 2287-4577 pISSN : 2287-9099

Research Paper
J Inf Sci Theory Pract 7(2): 23-31, 2019
https://doi.org/10.1633/JISTaP.2019.7.2.2

Impact of Open Access Models on Citation Metrics

Irina K. Razumova*
National Electronic Information Consortium,
Moscow, Russia
E-mail: razumova@neicon.ru

Alexander Kuznetsov
National Electronic Information Consortium,
Moscow, Russia
E-mail: kouz@neicon.ru

ABSTRACT
We report results of selection-bias-free approaches to the analysis of the impact of open access (OA) models on citation metrics. We studied reference groups of Gold and Green OA articles and the group of non-OA (Paywall) articles with the new functionality of the Web of Science Core Collection database, the InCites platform of Clarivate Analytics, and the Dimensions database of Digital Science. For each reference group we obtained the values of the percent of cited articles and citation impact and their dependence on the depth of the citation period. Different research fields were analyzed in two schemas of the InCites platform. We report the higher values and growth rates of the citation metrics: citation impact and %Cited, in the OA reference groups over the Paywall group. The Green OA articles demonstrate the highest values of citation metrics among all the OA models. Dependence of the value of citation impact on citation period follows linear law with R2 values close to 0.9–1.0. The overall annual growth rates of citation impact of the Green OA, Gold OA, and the Paywall articles, k equal, respectively, 3.6, 2.4, and 1.4 in Dimensions and 4.6, 3.6, and 2.3 in the Web of Science Core Collection. We suppose that earlier results reported for the articles in pure OA journals vs. articles in Paywall journals were affected by the high citation impact of the Green and Hybrid OA articles that could not be elucidated in the Paywall journals at that time.

Keywords: citation impact, %cited, open access, open access journals, hybrid journals, paywall

Open Access

Received date: March 09, 2019
Accepted date: May 23, 2019

***Corresponding Author:** Irina K. Razumova
Deputy Director of Science
National Electronic Information Consortium, Letnikovskaya St., 4/5, 115114,
Moscow, Russia
E-mail: razumova@neicon.ru

1. INTRODUCTION

1.1. Statement of the Problem

Results published in Zhu (2017) showed that 55 % of respondents in UK universities expected that open access (OA) articles would receive more citations. Only 8% of the respondents doubted that statement. Similar expectations were expressed by the respondents of the Russian survey conducted in 2018 (Razumova, Litvinova, Shvartsman, & Kuznetsov, 2018). To get the answer on the reality of such expectations in the new OA environment, we performed a study of the citation advantages of OA articles with the instruments and methodology recently developed in Web of Science Core Collection (WoS CC) and Dimensions.

1.2. Definitions

Following common definitions (Springer, 2019; Suber, 2006; Swan, 2012), we consider two main reference groups of open-access publications: Gold OA and Green OA. We assign to Gold OA the online journal articles published either in fully accessible OA (Pure-Gold-OA) journals or in Hybrid OA journals. Currently, the Directory of Open Access Journals (DOAJ) makes the world largest database of OA journals. Global citation indexes WoS CC, Scopus, and Dimensions use DOAJ as a source of Pure-Gold-OA articles. The hybrid journals are traditional subscription (Paywall) journals in which some of the articles are moved to OA (Hybrid OA articles). This stipulates the payment of an article processing charge (APC) to the publisher. Green OA refers to the author self-archiving a preprint or postprint versions of the article. Green OA articles are freely available for the general public on the websites of the institutional or subject-based OA repositories. For the purpose of this study, we will refer to the above reference groups of articles as Paywall, Gold OA, Pure-Gold-OA, Hybrid OA, and Green OA.

1.3. Literature Review

Since 2001 (Lawrence, 2001), many authors have reported the citation advantage of OA over non-OA or Paywall articles in separate research fields (Antelman, 2004; Eysenbach, 2006; Harnad et al., 2004; Kamat, 2018; Koler-Povh, Južnič, & Turk, 2014). In the first decade of the 21st century, Green OA citation impact was investigated and the OA citation advantage was confirmed (Metcalfe, 2005, 2006; Schwarz & Kennicutt, 2004; Wang, Liu, Mao, & Fang, 2015). However, a number of authors have argued this conclusion (Craig, Plume, McVeigh, Pringle, & Amin, 2007; Davis, Lewenstein, Simon, Booth, & Connolly, 2008; Davis & Walters, 2011).

No agreement has been reached yet and some authors have proposed a number of reasons for correlation between OA and citation advantage (Davis & Fromerth, 2007; Henneken et al., 2006; Kurtz et al., 2005; Moed, 2007).

The reasons were listed in Dorta-González, González-Betancor, & Dorta-González (2017), and can be summarized as follows:

(a) The OA postulate. Since OA articles are available for a wide audience, they get higher readership and citation.

(b) The early view postulate (Davis & Fromerth, 2007; Henneken et al., 2006; Kurtz et al., 2005; Moed, 2007). Green OA articles could be available online prior to their publication. They can, therefore, begin accumulating citations earlier than the paid access articles published at the same time and thus will have more citations because they have been available longer.

(c) The author selection bias postulate (Gaule & Maystre, 2011; McCabe & Snyder, 2014). Authors are more likely to provide OA to their highest quality articles, so OA articles will have more citations than paid-access articles.

(d) The APC selection bias. Average APC of about EUR 3,000 were listed among the barriers preventing authors from publishing articles in the OA model (Razumova et al., 2018; Zhu, 2017). Only rich and successful universities pay APCs for their authors. Successful universities perform high-quality research and thus provide the OA domain with high-quality articles that collect many citations.

(e) The grant selection bias. Currently, 108 out of the 148 largest research funders listed on the Sherpa Juliet website (Sherpa Juliett, 2019) require OA publishing (Gold OA) or OA archiving (Green OA) for articles supported by funder grants. The grants are issued to high-quality research, so this practice brings to OA high-quality articles.

Meanwhile, all the above-mentioned selection-bias postulates were formulated in the old OA environment. However, the OA world changes very quickly and during the last five years, the OA environment has changed dramatically. National OA policies and programs have been adopted that aim at 100% of publicly funded research to be published in Gold OA or Green OA. Many of them request or encourage Hybrid OA. Since 2014, OA policies have been launched in leading Western European countries: the United Kingdom, the Netherlands, France, Germany, Austria, Sweden, Denmark, Finland, etc. In July 2014, Higher Education Funding Council for England (HEFCE) announced the Research Excellence Framework (REF) 2021 OA policy: "The core of the REF 2021 OA policy is that journal articles and conference proceedings must be available

in an OA form to be eligible for the next REF. In practice, this means that these outputs must be uploaded to an institutional or subject repository" (HEFCE, 2014). The Plan S (2018) of the international consortium of research funders (cOAlition S, 2019) mandates that, from 2020, all articles funded by Plan S signatories will be published in compliant OA journals (Gold OA) or platforms (Green OA).

Upon realization of the national OA policies, new licenses were negotiated with the world leading publishers: Springer, Wiley, Elsevier, and so on. The list of newly negotiated licenses counts up to 15–20 publishers. The licenses include provisions that enable national corresponding authors to publish their articles in Gold (Hybrid) OA without paying APC.

The latest changes are resulting in the fast growth of the share of OA articles in the world publication flow. Thus, 28% of the 2016 articles in WoS CC were published in Green or Gold OA. Even higher numbers were reported in Science-Metrix (2018): As measured in Q3 2016, the percentage of OA articles in the WoS CC and science databases varied from 55% to 57% for the publication years within 2009–2014.

The new environment removes the OA selection bias. The author selection bias has no effect as OA publication becomes a requirement of funding bodies and authors are forced to publish all their articles in OA irrespective of their subjective choice. As APC are waived for authors, the APC barrier does not exist anymore.

2. MATERIALS AND METHODOLOGY

2.1. Citation Impact of Hybrid OA and Paywall Articles in Hybrid Journals of the Royal Society of Chemistry

2.1.1. The Gold-for-Gold Project of the Royal Society of Chemistry in Russia

In 2017-2018 together with the Royal Society of Chemistry (RSC), we analyzed the citation impact of Hybrid OA and Paywall articles of the Russian authors accepted for publication in the RSC hybrid journals. The articles were moved to the Hybrid OA mode within the Gold-for-Gold project of the RSC in Russia (G4G-RU).

The G4G-RU Project created a unique situation when 143 Russian articles accepted for publication in 2016 were moved to Hybrid OA almost at one time, namely in February to March 2017. The APC was waived for the Russian authors. This removed the problem of the APC barrier and the author choice bias, as all participating authors agreed to transform to OA all their articles compliant with the project.

We created two reference groups: the 143 OA articles (Hybrid OA) and 360 non-OA (Paywall) articles. All of the articles were published in the same hybrid journals presented in WoS CC: *Analyst, Analytical Methods, Catalysis Science & Technology, Chemical Communications, Crystengcomm, Dalton Transactions, Faraday Discussions, Green Chemistry, Journal of Materials Chemistry, Molecular Biosystems, Nanoscale, New Journal of Chemistry, Organic & Biomolecular Chemistry, Photochemical & Photobiological Sciences, Physical Chemistry Chemical Physics*, and *Soft Matter*. At the beginning of the project the journal *RSC Advances* was converted to the Pure-Gold-OA mode, so we excluded it from the list of analyzed hybrid journals.

Citation impact in each reference group was set to zero at the start of the April 2017 project. We controlled the number of citations of each article and of the package of articles as a whole. Citation impact values were calculated quarterly and the dependence of citation impact on the citation period was built within the first year after the start of the project.

2.1.2. Country-level Citation Impact of Hybrid OA and Paywall articles in RSC Hybrid Journals

In April 2018, in view of the new WoS CC OA functionality, we analyzed the country-level citation impact of the OA and Paywall articles published in 2016 in the whole domain of the RSC hybrid journals included in the WoS CC. Along with Russia, we selected two other countries which waived the RCS APCs for their corresponding authors: the United Kingdom and the Netherlands. The list of 40 RSC journals was obtained using the publisher filter of the InCites platform, the Pure-Gold-OA journals being excluded from the list. Searching for the 2016 articles was performed in WoS CC and Gold OA (Yes/No) filters were applied.

2.2. OA Models in WoS CC and Dimensions Databases

In this study, we analyzed the datasets of the 69 mln. article database WoS CC and the 96 mln. article Dimensions database. The filters used in WoS CC were: the journal article domain in Science Citation Index Expanded (SCIE), the Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (AHCI) and Emerging Sources Citation Index (ESCI), fixed publication year in the interval 2009–2017, OA (Yes/No), Gold OA (Bronze) (Yes/No), and Green OA (Yes/No). The filters used in Dimensions were: All publications or articles, Green OA (repositories), and Gold OA (journal publications).

We studied the reference datasets of the Green OA, Hybrid OA, Pure-Gold-OA, and Paywall articles. Newly developed

services of WoS CC and Dimensions enable article-level selection and analysis of the groups of Green OA, Gold OA, and All articles. Subtraction of Green OA and Gold OA data from the data on the All-article group gives the Paywall data. The Gold OA datasets in WoS CC and Dimensions include both the Hybrid OA and Pure-Gold-OA articles. The OA dataset of the InCites platform includes the Pure Gold OA articles from WoS CC only. To use the InCites instruments, we prepared the Gold OA, Green OA, and Paywall datasets in WoS CC and saved them to InCites. Subtraction of Pure-Gold-OA data obtained in InCites from the Gold OA data of WoS CC saved to InCites gives us the WoS CC Hybrid OA data exclusively.

2.2.1. Measured Citation Metrics

The reference datasets of the Green OA, Hybrid OA, Pure-Gold-OA, and Paywall articles were analyzed with the InCites functionality. The InCites instruments provide the following indicators:

- Number of articles
- Number of citations
- Citation impact
- %Cited
- Category Normalized Citation Impact

According to the InCites definition, the Category Normalized Citation Impact of a document is calculated by dividing the actual count of citing items by the expected citation rate for documents with the same document type, year of publication, and subject area.

In the Dimensions database, the reference datasets of Green OA, Gold OA, and Paywall articles were selected and studied with the Dimensions functionality. The indicators provided are as follows:

- Number of articles
- Number of citations
- Citation impact
- %Cited
- Field Citation Ratio (FCR)
- Relative Citation Ratio (RCR)

According to the definitions of the Dimensions database (Dimensions, 2019), the FCR is calculated by dividing the number of citations a paper has received by the average number received by documents published in the same year and in the same Fields of Research (FoR) category. The RCR is calculated for all PubMed publications which are at least two years old. It is calculated as the citations of a paper, normalized to the citations received by National Institutes of Health (NIH)-funded publications in the same area of research and year. The area of research is defined by the corpus of publications co-cited with the article of interest (the "co-citation network").

Datasets of the OA and Paywall articles of a given publication year were analyzed within the 2009–2017 interval, as in the 3rd quarter of 2018. We obtained dependence of citation impact and %Cited on the citation period. The latter is calculated as the number of years passed after the publication.

To investigate citation impact in separate research areas we performed the above analysis in major research areas in the schemas of the Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR) and Global Institutional Profiles Project (GIPP), available in InCites. Articles were grouped either in twelve major research fields of the ANVUR classification (Agricultural and Veterinary Sciences, Biology, Chemistry, Civic Engineering and Architecture, Earth Science, Economics and Statistics, Industrial and Information Engineering, Mathematics and Informatics, Medicine, Multidisciplinary (excluded), Physics, and Psychology) or in six major research fields of the GIPP classification: Arts & Humanities, Clinical, Pre-Clinical & Health, Engineering & Technology, Life Sciences, Physical Sciences, and Social Sciences.

To avoid the impact of grant selection bias, we studied the Green OA, Gold OA, and Paywall datasets of articles published with the grant support of the NIH that mandates Green and Gold OA publications.

To eliminate the impact of author selection bias and the APC barrier, we studied articles generated in countries having OA policy and author APCs waived: the United Kingdom, where APCs fees were paid within the block grants provided to the universities (UK Research and Innovation, 2013), and the Netherlands, where APCs were paid by universities upfront together with subscription fees (Openaccess.nl, 2019).

3. RESULTS AND DISCUSSION

3.1. Citation Impact of Hybrid OA and Paywall Articles in Hybrid Journals of the Royal Society of Chemistry

3.1.1. The Gold-for-Gold Project of the Royal Society of Chemistry in Russia

We studied the dependence of citation impact on citation period for two reference groups of the Russian articles published in the hybrid journals of RSC. Results are shown in Fig. 1.

In both cases, the y(x) dependence follows the linear law. In

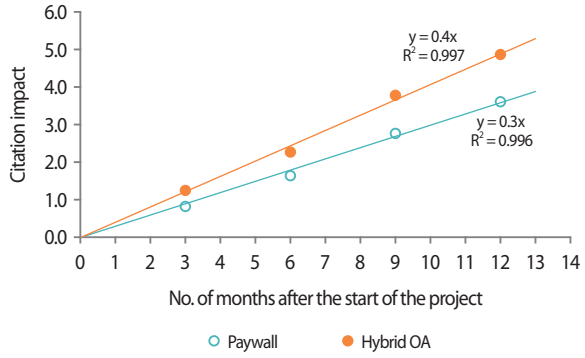


Fig. 1. Dependence of the citation impact on citation period for two groups of Hybrid open access (OA) and Paywall articles published in Royal Society of Chemistry hybrid journals in 2016. Measured in April 2018 in Web of Science Core Collection.

the case of the Hybrid OA articles: y=0.4x with R^2= 0.997. In the case of the Paywall articles, y=0.3x with R^2=0.996. Here, y is the citation impact and x is the citation period calculated in months from the start of the experiment in April 2017. If we count x in years, we will get y=4.8x and y=3.6x for OA and NON-OA articles, respectively. Here 4.8 and 3.6 are the growth rates of the Hybrid OA and Paywall citation impact within the first year after the start of the project.

3.1.2. Country-level Citation Impact of Hybrid OA and Paywall Articles in RSC Hybrid Journals

Results of the country-level analysis performed in April 2018 are shown in Table 1. The ratio of the citation impact of the Hybrid OA to that of the Paywall articles in the RSC journals is as follows: United Kingdom, 1.37; Netherlands, 1.33; and Russia, 1.36 (Table 1).

Thus we can conclude that the values of the citation impact of Russian articles measured in the domain of all hybrid RSC journals and those we obtained in the GFG-RU experiment are nearly the same. In the first year after publication, the Hybrid OA articles show a 35±2% citation advantage over Paywall articles published in the RSC hybrid journals.

3.2. Impact of OA Models on Citation Metrics of Articles in WoS CC and Dimensions Databases

3.2.1. Overall Values in All Subject Areas

In this section we report the overall analysis of citation metrics of the Dimensions articles published within 2009–2017. No bias-free filters were applied in these experiments. All the parameters were obtained in the 2nd quarter of 2018. The data set retrieved is the number of articles, the number of citations, %Cited, citation impact, RCR Mean, and FCR Mean (Table 2).

For the citation impact and %Cited we retrieved the data for OA and Paywall articles published in different publication years and built dependence of both parameters on the depth of the citation period.

The temporal dependence of %Cited values of the OA and Paywall articles in WoS CC and Dimensions is shown in Fig. 2.

The benchmark of the %Cited values of the Green OA articles in WoS CC, and Green OA, Gold OA, and Paywall articles in Dimensions are 99%, 89%, 74%, and 60%, respectively.

Table 1. Citation impact of Hybrid OA and Paywall articles published in hybrid Royal Society of Chemistry journals in 2016

Country	Hybrid OA model	Paywall model	Citation impact ratio (Hybrid OA to Paywall)
United Kingdom	7.34	5.37	1.37
Netherlands	6.95	5.23	1.33
Russia	4.9	3.6	1.36

Measured in April 2018 in Web of Science Core Collection.
OA, open access.

Table 2. Citation metrics of the Gold OA, Green OA, and Paywall articles of the Dimensions database published in 2009–2017

OA/non-OA article model	No. of publications	No. of citations	%Cited	Citation impact	RCR Mean	FCR Mean
Green OA	4,153,198	84,708,436	74	20	1.27	2.27
Gold OA	11,202,966	199,949,880	69	18	1.22	1.71
Paywall	80,694,884	712,281,155	47	9	0.66	1.1
All	96,051,048	996,939,471	50	10	0.75	1.32

Measured in second quarter 2018.
OA, open access; RCR, Relative Citation Ratio; FCR, Field Citation Ratio.

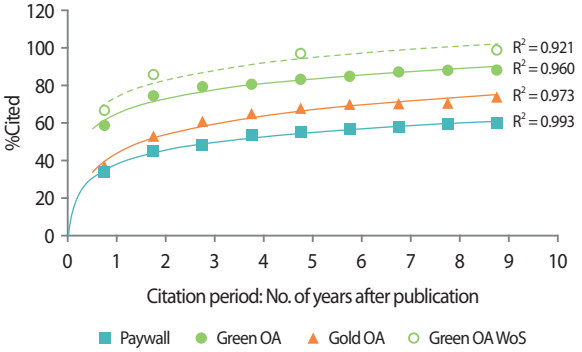


Fig. 2. Dependence of the %Cited values of the Green open access (OA), Gold OA, and Paywall articles on the depth of citation period. Measured in second quarter 2018 in Web of Science (WoS) Core Collection and Dimensions.

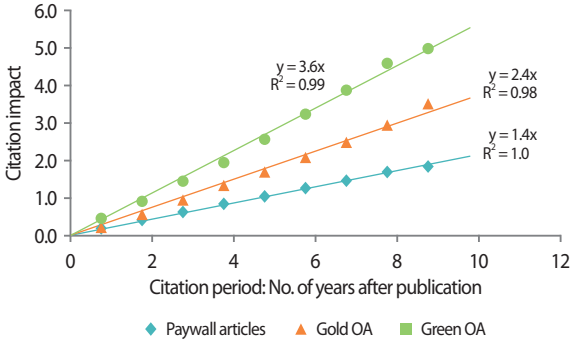


Fig. 3. Dependence of citation impact on citation period for Green open access (OA), Gold OA, and Paywall articles in the Dimensions database. Measured in second quarter 2018.

The dependence of citation impacts on the depth of the citation period for different OA models is shown in Fig. 3.

Within the nine-year period, temporal dynamics of the citation impact fit the linear dependence y=kx with R^2=0.99±0.01, where y is the citation impact, x is the number of years passed after the publication year, and k is the growth rate of the citation impact: k=3.6 for the Green OA, 2.4 for the Gold OA, and 1.4 for the Paywall articles.

Similar results were obtained for the WoS CC OA datasets processed with the InCites instruments. The k values for the WoS CC articles are 4.6, 3.6, and 2.3, respectively, for the Green OA, Gold OA, and Paywall articles.

3.2.2. Citation Impact in Separate Research Areas. Avoidance of Author Selection Bias and APC barrier

It was proposed that the high values of citation impact of the Green OA articles could be affected by a large number of articles in the fields of Medicine and Health published in PubMed because of the OA mandatory policy of the NIH. Articles in the field of Medical Sciences get higher citations than those in many other research fields and thus affect overall values. That is why we analyzed the citation impact of Green OA, Hybrid OA, Gold OA, Pure-Gold-OA, and Paywall articles in separate research areas of the ANVUR and GIPP schemas available in InCites. The OA and NON-OA datasets of articles were selected with the WoS CC functionality and saved to InCites. In each subject area, we retrieved OA/NON-OA citation impact within the 2009–2017 publication years and built dependence of the citation impact on the citation period.

We found that in both schemas, similarly to the results in Fig. 3, the dependence of the value of citation impact (y) on the citation period (x) can be approximated by the linear law y=kx with R^2 values close to 0.9–1.0.

Table 3. The growth rates (k) of the citation impact in the OA and non-OA groups of articles and the raw data of all Web of Science Documents and Times Cited in different research areas of the GIPP schema

Research area in GIPP schema	Green OA	Hybrid OA	Gold OA	Paywall	Web of Science Documents	Times cited	Citation impact
Life Sciences	9.7	5.7	5.1	3.3	197,649	2,122,309	10.7
Clinical, Pre-Clinical & Health	8.2	5.6	4.6	2.9	190,415	1,798,562	9.4
Physical Sciences	4.7	5.6	5.4	3.3	145,286	1,407,983	9.7
Engineering & Technology	3.9	4.3	3.9	2.8	104,222	720,625	6.9
Social Sciences	2.8	3.1	2.7	1.4	139,430	527,297	3.8
Arts & Humanities	0.9	1.1	0.9	0.3	40,781	35,847	0.9
Overall	6.4	5.2	4.7	2.5	633,980	4,963,537	7.8

The raw data in the 2014–2017 window.
OA, open access; GIPP, Global Institutional Profiles Project.

Table 4. The OA/Paywall ratio of the growth rates of citation impact in different research areas of the GIPP schema

Research area	Green OA/ Paywall	Hybrid OA/ Paywall	Gold OA/ Paywall
Life Sciences	2.9	1.7	1.5
Clinical, Pre-Clinical & Health	2.8	1.9	1.6
Physical Sciences	1.4	1.7	1.6
Engineering & Technology	1.4	1.5	1.4
Social Sciences	2.0	2.2	1.9
Arts & Humanities	3.0	3.7	3.0
Overall	2.6	2.1	1.9

OA, open access.

The GIPP Schema. In this case we studied articles published in the United Kingdom (England only) and the Netherlands to avoid author selection bias and the APC barrier. The k values calculated in each FoR are listed in Table 3. The raw data for all reference groups and for the whole dataset are given in the Supplemental Materials section. For reference purposes, in Table 3 we present the raw data for all the articles in the 2014–2017 citation window for each FoR.

The k-values in Table 3 clearly indicate that articles in the different OA reference datasets have an obvious citation advantage over the Paywall articles. The OA/Paywall ratios of the growth rates of citation impact in the GIPP schema FoRs are given in Table 4.

Noteworthy is that in none of our experiments did we detect any effect on citation impact of the early view of the Green OA articles. The ratio of Green OA/Paywall growth rates of citation impact varies from 1.4 for Physical Sciences and Engineering & Technology to 2.9–3.0 for Life Sciences and Arts & Humanities. The Hybrid OA/Paywall ratio varies from 1.5 in Engineering & Technology to 3.7 in Arts & Humanities. The Pure-Gold-OA data is not reported in this section as the temporal dependence of Pure-Gold-OA citation impact is not linear and the growth rate cannot be obtained. However, the Gold OA group that comprises both Hybrid OA and Pure-Gold-OA articles shows lower k values than those for Hybrid OA articles and indicates the lower citation impact of the Pure-Gold-OA articles. It would be of interest to study the behavior of this group in more detail.

Comparison of the Green OA group with the group of the Hybrid OA shows that the Green OA articles have higher citation impact in the research areas of Life Sciences and Clinical, and Pre-Clinical & Health. In Physical Sciences and Engineering & Technology, the citation impact of the Hybrid OA articles prevails.

The ANVUR Schema. In the case of the ANVUR schema, we investigated the overall dataset of the WoS CC articles published all over the world. Using the InCites instruments and the OA (Yes/No) filter, we could reconstruct results of Dorta-González et al. (2017) and elucidate the group of Pure-Gold-OA articles, that is, articles published exclusively in the Pure-Gold-OA journals and the group of Non-Pure-Gold-OA articles. In Dorta-González et al. (2017), those groups were defined as OA and non-OA articles. However, the non-Pure-Gold-OA group includes not only the Paywall but also the Green OA and Hybrid OA articles. As follows from Tables 3 and 4, the citation impact of the Green OA and Hybrid OA articles is much higher than that of the truly Paywall articles. Therefore, the citation impact of the group of non-Pure-Gold-OA articles is heavily affected by OA and its comparison with the citation impact of the Pure-Gold-OA articles is incorrect. Fig. 4 confirms that the comparison made using the InCites instruments leads to similar conclusions that were reported in Dorta-González et al. (2017).

The Green OA group of articles has the highest citation impact among all OA/non-OA groups in all research fields. Except for Biology, Earth Science, and Physics, the values of citation impact of the non-Pure-Gold-OA articles exceed those of the Pure-Gold-OA articles. This fits the results of Dorta-González et al. (2017), in which the authors found no citation advantage of OA articles. However, in our opinion this conclusion will be different if Hybrid OA and Green OA articles are considered. In our next study, we plan to double-check this statement.

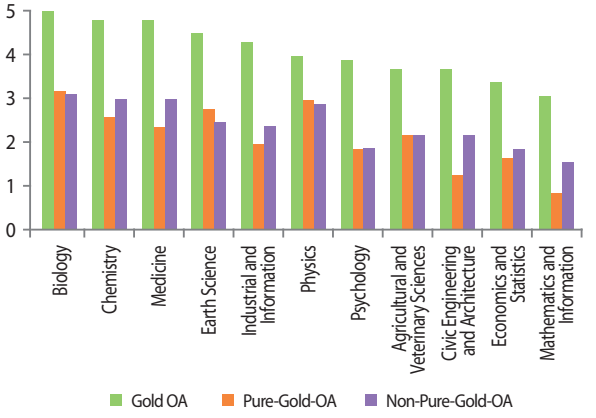


Fig. 4. Growth rates of citation impact in the groups of Green open access (OA), Pure-Gold-OA, and non-Pure-Gold-OA articles. The ANVUR (Italian National Agency for the Evaluation of Universities and Research Institutes) classification schema of subject areas. Web of Science Core Collection database and InCites platform. Retrieved in third quarter 2018.

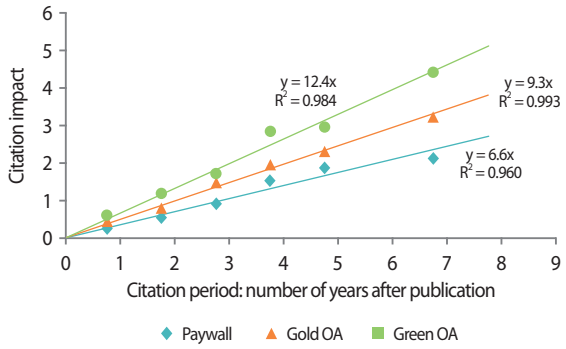


Fig. 5. Temporal dependence of citation impact in the groups of Green open access (OA), Gold OA, and Paywall articles published in the United Kingdom and the Netherlands with National Institutes of Health funding. Web of Science Core Collection database and InCites platform. Retrieved in third quarter 2018.

3.3. Eliminating Grant Selection Bias

As was indicated in Section 2.2, we investigated the citation impact of the Green OA, Gold OA, and Paywall articles published with the grant support of the National Institutes of Health that mandates Green OA and Gold OA publications. To eliminate author selection bias in this research, we filtered the WoS CC articles published in the United Kingdom and the Netherlands that mandate OA publishing.

We obtained dependence of citation impact of Green OA, Gold OA, and Paywall articles published in WoS CC on citation period, as in the 3rd quarter of 2018. Citation period was calculated as the number of years passed after the publication year. Publication year was selected within the 2009–2017 interval. Results are shown in Fig. 5.

Temporal dependence of citation impact follows the linear law y=kx in all three groups of articles. The growth rates of citation impact equal 12.4, 9.3, and 6.6 for the Green OA, Gold OA, and Paywall articles, respectively. Thus, the results obtained demonstrate the citation advantage of the OA articles.

4. CONCLUSION

We used several approaches to eliminate author selection bias, the APC barrier, and grant selection bias in studying the citation impact of different groups of OA and NON-OA articles in the WoS CC and Dimensions databases. Irrespectively of the bias filters, the results of this analysis ground the conclusion on the higher percent of cited articles and citation impact of OA as compared with Paywall articles. The Green OA articles demonstrate the highest values of citation metrics among all the OA models. The citation impact fits the linear dependence on the depth of the citation period: y=kx. Here, y is the citation impact, x is the depth of the citation period equal to the number of years passed after the publication year, and k is the growth rate of the citation impact. The growth rates of citation impact of the Green OA, Gold OA, and Hybrid OA articles exceed those of the Paywall articles. The values and the growth rates of citation impact vary in the different research areas. The detailed relevant studies of different research areas fall beyond the scope of the current article and could be the topic of future research.

No measurable effect of the early view postulate was detected for the citation impact of the Green OA articles. We also argue the earlier results reported no citation advantage for Pure-Gold-OA articles (articles published in pure-open-access journals) vs. non-Pure-Gold-OA articles (Dorta-González et al., 2017). In our opinion, the high level of the citation impact of non-Pure-Gold-OA articles measured in Dorta-González et al. (2017) was caused by the high citation impact of the Green OA and Hybrid OA articles that could not be eliminated in the Paywall journals at that time.

ACKNOWLEDGMENTS

The authors express their gratitude to Ivan Sterligov, Director of Scientometrics Center of National Research University Higher School of Economics, Moscow, Russia for valuable comments and fruitful ideas.

SUPPLEMENTAL MATERIALS

Supplemental materials are available from https://doi.org/10.1633/JISTaP.2019.7.2.2

REFERENCES

Antelman, K. (2004). Do open-access articles have a greater research impact?. *College & Research Libraries*, 65(5), 372-382.

cOAlition S. (2019). *About Plan S*. Retrieved Mar 11, 2019 from https://www.coalition-s.org.

Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics*, 1(3), 239-248.

Davis, P. M, & Walters, W. H. (2011). The impact of free access

to the scientific literature. A review of recent research. *Journal of the Medical Library Association*, 99(3), 208-217.

Davis, P. M., & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles?. *Scientometrics*, 71(2), 203-215.

Davis, P. M., Lewenstein, B. V, Simon, D. H., Booth, J. G., & Connolly, M. J. L. (2008). Open access publishing, article downloads, and citations: Randomised controlled trial. *BMJ*, 337, a568.

Dimensions. (2019). *Dimensions support*. Retrieved Mar 11, 2019 from https://dimensions.freshdesk.com/support/solutions/folders/23000029958.

Dorta-González, P., González-Betancor, S. M., & Dorta-González, M. I. (2017). Reconsidering the gold open access citation advantage postulate in a multidisciplinary context: An analysis of the subject categories in the Web of Science database 2009-2014. *Scientometrics*, 112(2), 877-901.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, 4(5), e157.

Gaule, P., & Maystre, N. (2011). Getting cited: Does open access help?. *Research Policy*, 40(10), 1332-1338.

Harnad, S., Brody, T., ValliÃ¨ res, F. O., Carr, L., Hitchcock, S., Gingras, Y., ... & Hilf, E. R. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30(4), 310-314.

HEFCE. (2014). *Higher Education Funding Council for England*. Retrieved Mar 11, 2019 from http://www.hefce.ac.uk.

Henneken, E. A., Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Thompson, D., & Murray, S. S. (2006). Effect of e-printing on citation rates in astronomy and physics. *Journal of Electronic Publishing*, 9(2). doi:10.3998/3336451.0009.202.

Kamat, P. (2018). Impact of open access papers in hybrid journals. *ACS Energy Letters*, 3, 410-411.

Koler-Povh, T., Južnič, P., & Turk, G. (2014). Impact of open access on citation of scholarly publications in the field of civil engineering. *Scientometrics*, 98(2), 1033-1045.

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S. S. (2005). The effect of use and access on citations. *Information Processing and Management*, 41(6), 1395-1402.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 411(6837), 521.

McCabe, M. J., & Snyder, C. M. (2014). Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry*, 52(4), 1284-1300.

Metcalfe, T. S. (2005). The rise and citation impact of astro-ph in major journals. *Bulletin of the American Astronomical Society*, 37, 555-557.

Metcalfe, T. S. (2006). The citation impact of digital preprint archives for solar physics papers. *Solar Physics*, 239(1-2), 549-553.

Moed, H. F. (2007). The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, 58(13), 2047-2054.

Openaccess.nl. (2019). *Openaccess.nl*. Retrieved Mar 11, 2019 from https://www.openaccess.nl/en.

Plan S. (2018). *From principles to implementation: cOAlition S releases implementation guidance on Plan S*. Retrieved Mar 11, 2019 from https://www.coalition-s.org/implementation.

Razumova, I. K., Litvinova, N. N., Shvartsman, M. E., & Kuznetsov, A. Y. (2018). Attitude to open access in Russian scholarly community: 2018. Survey results and analysis. *Scholarly Research and Information*, 1(1), 6-21.

Schwarz, G. J., & Kennicutt, R. C. (2004). Demographic and citation trends in astrophysical journal papers and preprints. *arXiv preprint*, arXiv:astro-ph/0411275.

Science-Metrix. (2018). *Analytical support for bibliometrics indicators: Open access availability of scientific publications*. Retrieved Mar 11, 2019 from http://www.science-metrix.com/sites/default/files/science-metrix/publications/science-metrix_open_access_availability_scientific_publications_report.pdf.

Sherpa Juliett. (2019). *Browse funders*. Retrieved Mar 11, 2019 from http://v2.sherpa.ac.uk/view/funder_list/1.html.

Springer. (2019). *What is open access?*. Retrieved Mar 11, 2019 from https://www.springer.com/gp/authors-editors/authorandreviewertutorials/open-access/what-is-open-access/10286522.

Suber, P. (2006). *Open access overview*. Retrieved Mar 11, 2019 from http://www.earlham.edu/~peters/fos/overview.htm.

Swan, A. (2012). *Policy guidelines for the development and promotion of open access*. Retrieved Mar 11, 2019 from https://unesdoc.unesco.org/ark:/48223/pf0000215863.

UK Research and Innovation. (2013). *Open access block grants*. Retrieved Mar 11, 2019 from https://www.ukri.org/funding/information-for-award-holders/open-access/open-access-policy/open-access-block-grants.

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics*, 103(2), 555-564.

Zhu, Y. (2017). Who supports open access publishing? Gender, discipline, seniority and other factors associated with academics' OA practice. *Scientometrics*, 111(2), 557-579.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice
eISSN : 2287-4577 pISSN : 2287-9099

Research Paper
J Inf Sci Theory Pract 7(2): 32-39, 2019
https://doi.org/10.1633/JISTaP.2019.7.2.3

Facial Data Visualization for Improved Deep Learning Based Emotion Recognition

Seung Ho Lee*

Department of Future Technology, Korea University of Technology and Education, Cheonan, Korea
E-mail: leesh903@koreatech.ac.kr

ABSTRACT

A convolutional neural network (CNN) has been widely used in facial expression recognition (FER) because it can automatically learn discriminative appearance features from an expression image. To make full use of its discriminating capability, this paper suggests a simple but effective method for CNN based FER. Specifically, instead of an original expression image that contains facial appearance only, the expression image with facial geometry visualization is used as input to CNN. In this way, geometric and appearance features could be simultaneously learned, making CNN more discriminative for FER. A simple CNN extension is also presented in this paper, aiming to utilize geometric expression change derived from an expression image sequence. Experimental results on two public datasets (CK+ and MMI) show that CNN using facial geometry visualization clearly outperforms the conventional CNN using facial appearance only.

Keywords: facial expression recognition, convolutional neural network, facial landmark points, facial geometry visualization

Open Access

Received date: March 05, 2019
Accepted date: May 29, 2019

***Corresponding Author:** Seung Ho Lee
Assistant Professor
Korea University of Technology and Education, 1600 Chungjeol-ro, Byeongcheon-myeon, Dongnam-gu, Cheonan 31253, Korea
E-mail: leesh903@koreatech.ac.kr

1. INTRODUCTION

Data science has emerged as a core area of information science in recent years. Data science uses scientific algorithms or systems to extract knowledge and insights from structured and unstructured data. Particularly, recent advances in computer vision and machine learning (e.g., deep learning) techniques can successfully reduce costs of analyzing large-scale visual data (e.g., image or video), which is one of the representative forms of unstructured data. A convolutional neural network (CNN) has been most successfully used among deep learning methods for visual data.

Face analysis can play an important role in visual data mining. Automatic face recognition (or face identification) is a mature and widely used approach, which allows face images to be organized by the identities of persons. For example, personal photos in social network services can be automatically labeled with the names of persons, which is called name tagging (Choi, Neve, Plataniotis, & Ro, 2011). Facial expression recognition (FER) has been increasingly important due to its emerging applications. Facial expression is one of the most natural and powerful tools for non-verbal human communications (Sandbach, Zafeiriou, Pantic, & Yin, 2012), which conveys emotional states, such as surprise, intention, and interest. Leveraging expressive signals would have many practical applications. For example, bunches of emotional response data of customers' faces to online video adverts can be unobtrusively collected (via smart phone camera or web camera) and analyzed for market research purposes. In addition, for improving road safety, driver state monitoring can be adopted in automobiles, which understands the driver's emotional states (e.g., anger that can negatively affect his/her driving) and moods in real time. Considering the practical importance, this paper focuses on improving deep learning (i.e., CNN) to more accurately predict the emotional state from an expression image.

A number of traditional FER methods extract hand-crafted appearance features to capture pixel intensity changes of facial expression images. These features include local binary patterns (Huang, Wang, & Ying, 2010), local phase quantization (Wang & Ying, 2012), two-dimensional principal component analysis (Yang, Zhang, Frangi, & Yang, 2004), color texture features (Lee, Kim, Ro, & Plataniotis, 2013), and so on. To classify expression features, many well-known classifiers such as the support vector machine (Bartlett et al., 2005) or sparse representation classifier (Wright, Yang, Ganesh, Sastry, & Ma, 2009) have been employed and evaluated.

More recently, due to dramatically increasing processing ability (e.g., Graphics Processing Unit processing), more powerful FER methods have been proposed based on deep learning techniques. Among various deep learning techniques, CNN (Lecun, Bottou, Bengio, & Haffner, 1998) has been perfectly designed to take expression image data as input and learn discriminative appearance features from the data.

Convolution layers in CNN automatically learn appearance features for FER. Specifically, during the training stage, the weight values of spatial filters (or kernels) are learned according to the input expression image and its ground truth (i.e., emotion class label). In the test stage, two dimensional feature maps (or activation maps) can be obtained by applying the learned spatial filters. These feature data are passed to fully connected layers in order to predict an emotion class of the input expression image.

Despite successful use of CNN based FER, one limitation is that most CNN based FER methods take the original expression image as input and extract appearance features only. However, it should be noted that geometric features are also important for recognizing facial expression, as pointed out in Chen et al. (2012). For extracting geometric features, landmark points or feature points need to be detected and processed (Kotsia & Pitas, 2007). Jung, Lee, Yim, Park, & Kim (2015) have attempted to incorporate geometric information into CNN based FER. In the method in Jung et al. (2015), while CNN takes expression images, another deep neural network takes the xy coordinates of the landmark points. The outputs of the two different deep neural networks are combined by the fine-tuning method proposed in Jung et al. (2015). However, this approach has difficulty analyzing the spatial relation between two-dimensional facial appearance (i.e., expression image) and one-dimensional facial geometry (i.e., landmark points). As a result, the effectiveness of fusing geometric and appearance information could be limited due to the different domains of analysis.

To make full use of the discriminating capability of CNN, this paper suggests incorporation of facial geometry visualization. Instead of an original expression image, the expression image with visualizations of landmark detection and processing is used for CNN based FER. The visualizations on an expression image have been experimentally investigated in this paper as to whether the process is suitable for improving the discriminating power of CNN. Note that facial expression change appears in a continuous video sequence (Lee & Ro, 2016), which is a usual input format to realistic FER applications. In order to derive facial dynamics from the sequence, a simple CNN extension is also presented in this paper. More specifically, CNN takes two channel input, i.e., an expressive image with the facial geometry visualization and a non-expressive (or neutral) image with the facial geometry visualization. Experimental results on two public datasets, CK+ (Lucey et al., 2010) and MMI (Pantic,

Valstar, Rademaker, & Maat, 2005), show that CNN using facial geometry visualization clearly outperforms the conventional CNN using only facial appearance in terms of recognition accuracy. About up to 5% of improvement has been achieved for the case of one channel input (i.e., an expression image). From the CNN extension to two channel input, additional improvement has been achieved. In addition, it has also been demonstrated that the proposed method (i.e., CNN based FER using facial geometry visualization) can be comparable with some recent advances in recognition accuracy.

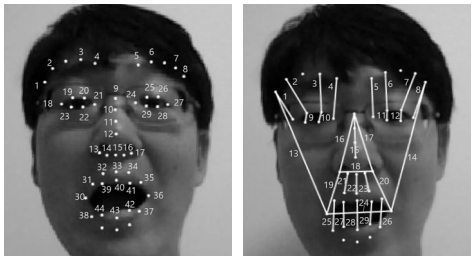
The rest of this paper is summarized as follows. Section 2 presents details about the proposed method. Section 3 presents experimental results to demonstrate the effectiveness of the proposed method. Conclusions are drawn in Section 4.

2. PROPOSED METHOD

This section describes a method to improve the discriminating power of CNN based FER through facial geometry visualization. Section 2.1 presents the method for static expression images. A simple extension for expression image sequences is given in Section 2.2.

2.1. Discriminative CNN with Facial Geometry Visualization

This section focuses on describing how to make use of facial geometric information for CNN based FER. From a grayscale expression image of N×N pixels, 49 landmark points are extracted (Fig. 1a). For the automatic landmark detection, the method in Asthana, Zafeiriou, Cheng, & Pantic (2014) is adopted. To make the facial geometry more meaningful for FER, two landmark points are connected with a line as shown in Fig. 1b. For the connections, 44 out of 49 landmark points are used and a total of 29 connecting lines can be visualized. As explained in Lee and Ro (2016), the set of 29 connecting lines could be closely related to facial muscles. For example, the lengths of the connecting line No. 1-8 represent the distances between the brows and the corresponding eyelids. Hence, these connecting lines are related to Action Unit (AU) 1 (inner brow raiser), AU5 (upper lid raiser), AU7 (lid tightener), and so on (Lee & Ro, 2016).



(a) (b)

Fig. 1. Facial geometry visualization with landmark points. (a) Forty-nine landmark points detected. (b) Twenty-nine kinds of lines connecting two landmark points.

Fig. 2 illustrates 29 connecting lines for six basic emotion classes (i.e., Anger, Disgust, Fear, Happy, Sad, and Surprise). One can see that the facial geometry visualizations (each of which is formed with the 29 connecting lines) look clearly different across the six emotion classes. It is expected that spatial relations between facial appearance and facial geometry could extract useful features for classifying emotional facial expressions. Therefore, the expression image with visualization of the 29 connecting lines is directly fed into CNN.

Fig. 3 illustrates the CNN model used in this paper. The processed expression image with facial geometry visualization is resized to 32×32 pixels. The first convolution layer employs 64 filters with sizes of 5×5. The second and third convolution layers employ 128 filters with sizes of 3×3, respectively. The three convolution layers produce 64, 128, and 128 feature maps, respectively. Next, use of a 2×2 max pooling layer is followed to reduce the spatial size of the feature maps and the computational cost of the network. Similar to conventional CNNs, a fully connected layer is included at the end of the network to classify emotion class. After the max pooling layer, each of the two-dimensional feature maps is converted into the one-dimensional feature maps which are suitable for the input to the fully connected layer. The output layer has six nodes (for six emotion classes in MMI) or seven nodes (for seven emotion classes in CK+). Through the output layer, an emotion class is predicted by finding the highest probabilistic score.

Anger Disgust Fear Happy Sad Surprise

Fig. 2. Facial geometry visualization defined by 29 connecting lines.

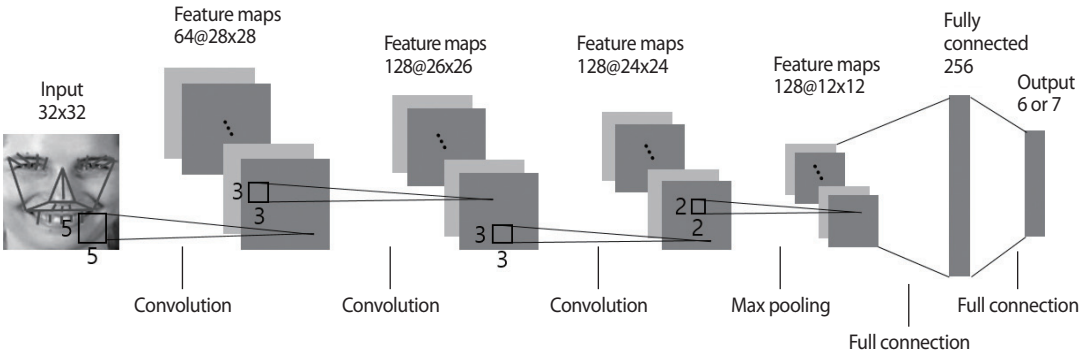


Fig. 3. Convolutional neural network model used in this paper.

2.2. Extension to Two Channel Input for Expression Image Sequence

In Section 2.1, CNN using one channel input is presented to classify a static expression image. Note that we may encounter sequences of facial expression where the face evolves from a neutral state to an emotional expressive state (Kotsia & Pitas, 2007). As extensively studied in several works (Chen et al., 2012; Kotsia & Pitas, 2007; Zafeiriou & Petrou, 2010; Donato, Stewart, Hager, Ekman, & Sejnowski, 1999), using both expressive state and neutral state can be useful to capture the expression change of a person. In this section, a simple CNN extension is described, which aims to capture more discriminative expression features from an expression image sequence.

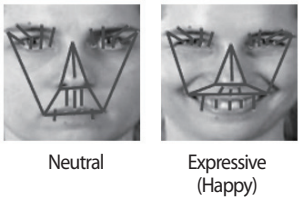


Fig. 4. Facial geometry visualizations for neutral image and expressive image.

Note that CNN has been basically designed to take a multichannel image as input. For example, CNN can take an RGB color image and learn discriminative features from three complementary color channels. Taking the structure of CNN into consideration, two channel input of the non-expressive (or neutral) image and expressive image is applied to the CNN model in Fig. 3. The neutral and expressive images are processed to include visualization of 29 connecting lines in a similar way to Section 2.1. The visualization result is shown in Fig. 4. The difference between the two is able to encode expression change information. An extension of CNN to two channel input is illustrated in Fig. 5. From the input and the first convolutional layer in Fig. 5, one can expect that appearance and geometry changes are learned considering both neutral and expressive states of a face. It should be noted that, for consistency, two channel input of neutral and expressive images is used in both the training stage and test stage.

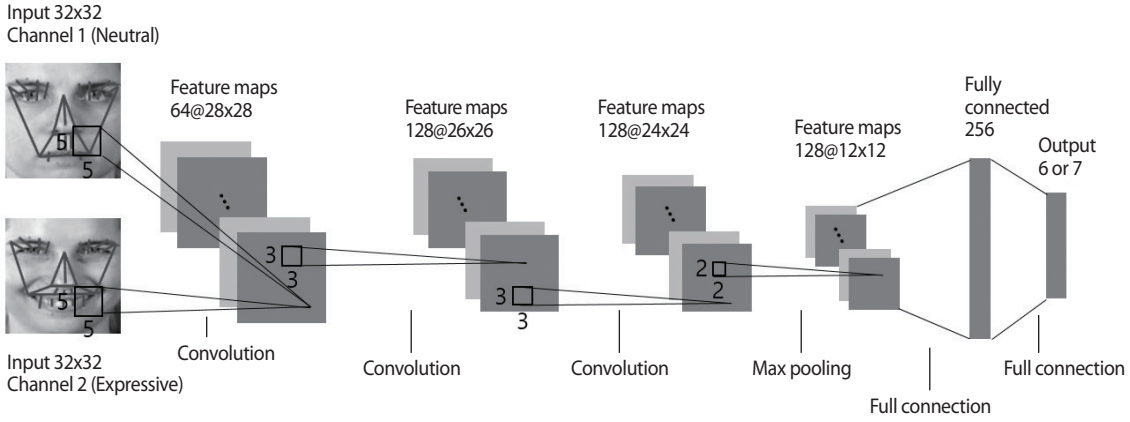


Fig. 5. Convolutional neural network extension to two channel input.

3. EXPERIMENT

3.1. Experimental Setup

The proposed method was evaluated with two public datasets, i.e., CK+ and MMI. Example expression images from the datasets are shown in Fig. 6. The dataset constructions for the experiment were performed as follows:

1. CK+ (Lucey et al., 2010): CK+ consists of 593 image sequences from 123 persons. From these, 325 video sequences of 118 persons were selected, which satisfied the criteria for one of the seven emotion classes. The selected 325 video sequences consisted of 45, 18, 58, 22, 69, 28, and 82 video sequences of Angry, Contempt, Disgust, Fear, Happy, Sadness, and Surprise, respectively. Ten-fold cross validation was used to measure recognition accuracy.
2. MMI (Pantic et al., 2005): 205 video sequences were collected from 30 persons. The dataset consisted of 31, 31, 27, 43, 32, and 41 video sequences of Anger, Disgust, Fear, Happy, Sad, and Surprise, respectively. Ten-fold cross validation was also used for this dataset.

For both datasets, from each video sequence, a pair of neutral frame and peak expression frames (i.e., most expressive frame) was manually selected by the author. From each frame, facial region was detected by using Viola Jones algorithm (Viola & Jones, 2004). The detected facial region was aligned based on two eye locations and cropped resulting in an expression image. To compute two eye locations, 49 facial landmark points were detected using the method in Asthana et al. (2014). The coordinates of the left eye and the right eye were obtained by averaging those of the facial landmark no. 18-23 and no. 24-29 (Fig. 1a), respectively. Forty-four out of 49 landmark points were used to visualize facial geometry with 29 connecting lines. The expression images with visualization of 29 connecting lines were resized to 32×32 pixels.

The CNN models in Fig. 3 and Fig. 5 were implemented using Python ver. 3.5.4 and an open-source deep learning library called Keras. As activation functions, Rectified Linear Unit (ReLU) and Softmax were adopted for the convolutional layers and the output layer, respectively. In order to learn the CNN model, the number of epochs and batch size were set to 45 and 30, respectively. In addition, Adam was selected as optimizer and its learning rate was set to 0.3.

3.2. Experimental Results

In this section, the effectiveness of using facial geometry visualization (i.e., visualization of 29 connecting lines on expression image) was investigated. For this purpose, four different FER methods were defined as follows:

1. Peak without facial geometry: An expression image with the highest expression intensity in each video sequence was used as input for CNN based FER. The expression image did not include the visualization of 29 connecting lines.
2. Peak with facial geometry (proposed): An expression image with the highest expression intensity in each video sequence was used as input for CNN based FER. The expression image included the visualization of 29 connecting lines.
3. Peak/neutral without facial geometry: A pair of peak expression image and neutral expression image was used as input for CNN based FER. The expression images did not include the visualizations of 29 connecting lines.
4. Peak/neutral with facial geometry (proposed): A pair of peak expression image and neutral expression image was used as input for CNN based FER. The expression images included the visualizations of 29 connecting lines.

Table 1 shows comparisons of the four FER methods on CK+. From the comparison results, two observations can be made. First, compared to using only a peak expression image, using a pair of peak expression and neutral expression images yields improved recognition accuracies. This is mainly because a neutral state can be useful to capture the pure expression change (neutral to expressive) present in an expression image sequence. Second, regardless of using a neutral image, including the visualization of 29 connecting lines on an expression image is clearly better in recognition accuracy than using the expression image without the visualization. By incorporating facial geometry visualization, about 3% to 5% of improvements are achieved for CNN based FER. Table 2 shows the confusion matrix for 'Peak/Neutral with facial geometry' on CK+. It is shown that Fear and Sad expressions are often misclassified as Surprise and Angry, respectively.

Table 3 shows comparisons results of the four FER methods on MMI. Note that MMI is much more difficult to analyze than CK+ for the following reasons. First, different persons make the same emotional expression differently as shown in Fig. 6b. Second, some persons have accessories such as glasses or head cloths (Fig. 6b). Although the average recognition accuracy in Table 3 is relatively low (70.47%), visible improvements are made by using facial geometry visualization, while achieving up to 74.11%. Table 4 shows the confusion matrix for 'peak/neutral with facial geometry' on MMI. It is observed that recognizing Fear is very difficult as it is often confused with Surprise, similar to the case in CK+. It is also observed that Anger and Disgust are confused with each other.

Fig. 6. Datasets used in experiment. (a) CK+. (b) MMI.

Table 1. Comparison results on CK+

Facial expression recognition method	Recognition rate (%)
Peak without facial geometry	85.87
Peak with facial geometry	88.34
Peak/neutral without facial geometry	87.74
Peak/neutral with facial geometry	92.63

Table 2. Confusion matrix on CK+ obtained by using 'peak/neutral with facial geometry'

Actual	Predicted						
	Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise
Angry	**91.11**	0.00	4.44	2.22	0.00	2.22	0.00
Contempt	5.56	**88.89**	0.00	0.00	0.00	5.56	0.00
Disgust	1.72	0.00	**96.55**	0.00	1.72	0.00	0.00
Fear	0.00	0.00	0.00	**76.00**	4.00	8.00	12.00
Happy	0.00	0.00	0.00	0.00	**100.00**	0.00	0.00
Sad	10.71	0.00	0.00	7.14	0.00	**82.14**	0.00
Surprise	1.22	0.00	0.00	1.22	0.00	3.66	**93.90**

Table 3. Comparison results on MMI

Facial expression recognition method	Recognition rate (%)
Peak without facial geometry	67.77
Peak with facial geometry	69.29
Peak/neutral without facial geometry	70.72
Peak/neutral with facial geometry	74.11

Table 4. Confusion matrix on MMI obtained by using 'Peak/neutral with facial geometry'

Actual	Predicted					
	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**67.74**	19.35	3.23	0.00	6.45	3.23
Disgust	19.35	**58.06**	6.45	6.45	6.45	3.23
Fear	7.41	3.70	**55.56**	3.70	7.41	22.22
Happy	4.65	0.00	2.33	**90.70**	2.33	0.00
Sad	9.38	6.25	6.25	3.13	**71.88**	3.13
Surprise	4.88	0.00	7.32	0.00	0.00	**87.80**

Table 5. Comparisons with recent advances in FER on MMI

FER method	Cross validation	Recognition rate (%)
Proposed	10-fold	74.11
GAN (Yang et al., 2018)	10-fold	73.23
Identity-Aware CNN (Meng et al., 2017)	10-fold	71.55
Partial matching (Lee & Ro, 2016)	10-fold	72.10
DTAGN-Joint (Jung et al., 2015)	10-fold	70.24
HMM (Weng, Wang, & Ji, 2013)	15-fold	51.50
ITBN (Weng et al., 2013)	15-fold	59.70

FER, acial expression recognition; GAN, generative adversarial network; CNN, convolutional neural network; DTAGN, deep temporal appearance-geometry network; HMM, hidden Markov model; ITBN, interval temporal Bayesian network.

3.3. Comparisons with Recent Advances

In this section, the proposed method (CNN based FER using facial geometry visualization) was compared with some recent advances in FER on MMI dataset under similar testing protocols (e.g., recognition of six emotion classes under 10- or 15-fold cross validation). Table 5 shows the comparison results. It is observed that the deep learning based methods (Jung et al., 2015; Yang, Ciftci, & Yin, 2018; Meng, Liu, Cai, Han, & Tong, 2017) achieve relatively high recognition accuracies ranging from 70.24 to 74.11. It should be noted that the proposed method is comparable with the DTAGN-Joint method (Jung et al., 2015) that also uses geometric information for deep neural networks. In the method (Jung et al., 2015), the detected facial landmark points are converted into one-dimensional data and fed into a deep neural network (fully connected layer). On the other hand, the proposed method directly processes the facial geometry and its original appearance as a single two-dimensional image data without any data conversion such as vectorization. Thus, the features learned via the convolutional layers could be two-dimensional and thus are more straightforward for visual emotion classification.

4. CONCLUSION

The ability to accurately recognize and interpret a person's facial expressions is a key to practical data science applications. For FER, CNN has been widely adopted because it can automatically learn discriminative appearance features from an expression image. For more accurate FER, this paper proposes a simple but effective method to incorporate geometric information into CNN. In the proposed method, instead of an original expression image that contains facial appearance only, the expression image with facial geometry visualization is used as input to CNN. Spatial relation between facial appearance and facial geometry could make the learned expression features more discriminative. For future work, the various visualization methods of facial geometry will be studied for further improving the proposed approach.

Note that the proposed method is very simple and easy to implement because it does not need to change the structure of conventional CNN models. Thus, it is believed that it could be practically used for various applications including emotion mining from online video adverts or driver state monitoring for road safety.

ACKNOWLEDGMENTS

This paper was partially supported by the New Professor Research Program of KOREATECH in 2018.

REFERENCES

Asthana, A., Zafeiriou, S., Cheng, S., & Pantic, M. (2014). Incremental face alignment in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23-28 June (pp. 1859-1866).

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005). Recognizing facial expression: Machine learning and application to spontaneous behavior. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 20-25 June (pp. 568-573). Piscataway: IEEE.

Chen, J., Chen, D., Gong, Y., Yu, M., Zhang, K., & Wang, L. (2012). Facial expression recognition using geometric and

appearance features. In *Proceedings of the 4th International Conference Internet Multimedia Computing and Service*, Wuhan, China (pp. 29-33). New York: ACM.

Choi, J. Y., Neve, W. D., Plataniotis, K. N., & Ro, Y. M. (2011). Collaborative face recognition for improved face annotation in personal photo collections shared on online social networks. *IEEE Transactions on Multimedia*, 13(1), 14-28.

Donato, G., Stewart, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10), 974-989.

Huang, M. W., Wang, Z. W., & Ying, Z. L. (2010). A new method for facial expression recognition based on sparse representation plus LBP. In *2010 3rd International Congress on Image and Signal Processing*, Yantai, China, 16-18 October (Vol. 4, pp. 1750-1754). Piscataway: IEEE.

Jung, H., Lee, S., Yim, J., Park, S., & Kim, J. (2015). Joint fine-tuning in deep neural networks for facial expression recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, 7-13 December (pp. 2983-2991).

Kotsia, I., & Pitas, I. (2007). Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Transactions on Image Processing*, 16(1), 172-187.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.

Lee, S. H., Kim, H., Ro, Y. M., & Plataniotis, K. N. (2013). Using color texture sparsity for facial expression recognition. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Shanghai, China, 22-26 April (pp. 1-6). Piscataway: IEEE.

Lee, S. H., & Ro, Y. M. (2016). Partial matching of facial expression sequence using over-complete transition dictionary for emotion recognition. *IEEE Transactions on Affective Computing*, 7(4), 389-408.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, San Francisco, CA, USA, 13-18 June (pp. 94-101). Piscataway: IEEE.

Meng, Z., Liu, P., Cai, J., Han, S., & Tong, Y. (2017). Identity-aware convolutional neural network for facial expression recognition. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition*, Washington, DC, USA, 30 May-3 June (pp. 558-565). Piscataway: IEEE.

Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005). Web-based database for facial expression analysis. In *2005 IEEE International Conference on Multimedia and Expo*, Amsterdam, Netherlands, 6 July (p. 5). Piscataway: IEEE.

Sandbach, G., Zafeiriou, S., Pantic, M., & Yin, L. (2012). Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 30(10), 683-697.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 137-154.

Wang, Z., & Ying, Z. (2012). Facial expression recognition based on local phase quantization and sparse representation. In *2012 8th International Conference on Natural Computation*, Chongqing, China, 29-31 May (pp. 222-225). Piscataway: IEEE.

Weng, Z., Wang, S., & Ji, Q. (2013). Capturing complex spatio-temporal relations among facial muscles for facial expression recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, 23-28 June (pp. 3422-3429).

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 210-227.

Yang, H., Ciftci, U., & Yin, L. (2018). Facial expression recognition by de-expression residue learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18-23 June (pp. 2168-2177).

Yang, J., Zhang, D. D., Frangi, A. F., & Yang, J. Y. (2004). Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1), 131-137.

Zafeiriou, S., & Petrou, M. (2010). Sparse representation for facial expressions recognition via l1 optimization. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, San Francisco, CA, USA, 13-18 June (pp. 32-39). Piscataway: IEEE.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice
eISSN : 2287-4577 pISSN : 2287-9099

Research Paper
J Inf Sci Theory Pract 7(2): 40-53, 2019
https://doi.org/10.1633/JISTaP.2019.7.2.4

Fake News in Social Media: Bad Algorithms or Biased Users?

Franziska Zimmer
Department of Information Science, Heinrich Heine University, Düsseldorf, Germany
E-mail: franziska.zimmer@hhu.de

Katrin Scheibe
Department of Information Science, Heinrich Heine University, Düsseldorf, Germany
E-mail: katrin.scheibe@hhu.de

Mechtild Stock
Stock-Kerpen, Kerpen, Germany
E-mail: mechtildstock@gmail.com

Wolfgang G. Stock*
Department of Information Science, Heinrich Heine University, Düsseldorf, Germany
E-mail: stock@phil.hhu.de

ABSTRACT
Although fake news has been present in human history at any time, nowadays, with social media, deceptive information has a stronger effect on society than before. This article answers two research questions, namely (1) Is the dissemination of fake news supported by machines through the automatic construction of filter bubbles, and (2) Are echo chambers of fake news man-made, and if yes, what are the information behavior patterns of those individuals reacting to fake news? We discuss the role of filter bubbles by analyzing social media's ranking and results' presentation algorithms. To understand the roles of individuals in the process of making and cultivating echo chambers, we empirically study the effects of fake news on the information behavior of the audience, while working with a case study, applying quantitative and qualitative content analysis of online comments and replies (on a blog and on Reddit). Indeed, we found hints on filter bubbles; however, they are fed by the users' information behavior and only amplify users' behavioral patterns. Reading fake news and eventually drafting a comment or a reply may be the result of users' selective exposure to information leading to a confirmation bias; i.e. users prefer news (including fake news) fitting their pre-existing opinions. However, it is not possible to explain all information behavior patterns following fake news with the theory of selective exposure, but with a variety of further individual cognitive structures, such as non-argumentative or off-topic behavior, denial, moral outrage, meta-comments, insults, satire, and creation of a new rumor.

Keywords: fake news, truth, information behavior, social media, filter bubble, echo chamber

Open Access

Received date: April 18, 2019
Accepted date: June 05, 2019

***Corresponding Author:** Wolfgang G. Stock
Professor
Department of Information Science, Heinrich Heine University, Universitätsstraße 1 D-40225 Düsseldorf, Germany
E-mail: stock@phil.hhu.de

1. INTRODUCTION

It is a truism that false propositions or even deceptions reach their recipients every day and everywhere. Fake news on online press sites and on social media is no exception. However, deceptive information "has had dramatic effect on our society in recent years" (Volkova & Jang, 2018, p. 575). Deceptions and fake news may possibly survive very well in environments of all kinds of social media, be it weblogs, microblogging services, social live streaming platforms, image and video sharing services, or social networking services. "Despite optimistic talk about 'collective intelligence,' the Web has helped create an echo chamber where misinformation thrives. Indeed, the viral spread of hoaxes, conspiracy theories, and other false or baseless information online is one of the most disturbing social trends of the early 21st century" (Quattrociocchi, 2017, p. 60), leading even to the "emergence of a post-truth world" (Lewandowsky, Ecker, & Cook, 2017, p. 357). Especially, such historically relevant events as the UK's Brexit vote (Bastos, Mercea, & Baronchelli, 2018), the 2016 presidential election in the United States (Allcott & Gentskow, 2017), and the excessive use of the term "fake news" by Donald Trump has led to discussions about the role of fake news in society. The related term "post-truth" was named word of the year for 2016 by the Oxford Dictionaries (2016).

In *The Guardian*, we read "social media filter bubbles and algorithms influence the election" in Great Britain (Hern, 2017). Similarly, for the Observer, "the problem isn't fake news, it's bad algorithms" (Holmes, 2016). The University of Amsterdam's *Master of Media* blog addresses filter bubbles as algorithms customizing our access to information (Mans, 2016). These three examples clearly demonstrate what the cause of fake news dissemination is: It is bad algorithms. Nevertheless, one may find divergent opinions in the popular press. The *New Statesman* claims, "Forget fake news of Facebook: the real filter bubble is you" (Self, 2016). Now, the cause of fake news distribution is the misleading information behavior of individual people, i.e. biased users. As filter bubbles and echo chambers are often discussed in the press Bruns (2019) asks, "are filter bubbles real," and are they overstated?

"Bad algorithms" are related to "filter bubbles," being applications of personalized information retrieval as well as of recommender systems. They lead the users to receive only an excerpt of (maybe false) propositions instead of the entire spectrum of appropriate information. A source for concrete algorithmic recommendations is the user's former information behavior, which is recognized by the machine. On the other hand, "bad user behavior" or "biased users" (Vydiswaran, Zhai, Roth, & Pirolli, 2012) refer to "echo chambers," which are loosely connected clusters of users with similar ideologies or interests, whose members notice and share only information appropriate to their common interests. The information behavior of the user in question in combination with other users' behaviors (e.g., commenting on posts or replying to comments) exhibits special patterns which may lead to the echo chamber effect (Bruns, 2017).

2. RESEARCH OUTLINE

First of all, the main concepts must be defined. *Fake news* is information including "phony news stories maliciously spread by outlets that mimic legitimate news sources" (Torres, Gerhart, & Negahban, 2018, p. 3977); it is misinformation (transmitting untrue propositions, nonconsidering the cognitive state of the sender) and disinformation (again, transmitting untrue propositions, but now consciously by the sender) (Shin, Jian, Driscoll, & Bar, 2018). Deception is a kind of disinformation which brings an advantage to the sender. Other authors compare fake news to satire and parody, fabrication, manipulation, and propaganda (Tandoc Jr., Lim, & Ling, 2018). The users' appraisement of a news story as fake or non-fake depends on the content of the story and—a little bit more—on the source of the transmitted information (Zimmer & Reich, 2018) as well as on the presentation format (Kim & Dennis, 2018).

This paper follows the well-known definition of *social media* by Kaplan and Haenlein (2010, p. 61): "Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content." Social Media includes, among other systems, weblogs, social networking services (such as Facebook), news aggregators (such as Reddit), knowledge bases (such as Wikipedia), sharing services for videos and images (such as YouTube and Instagram), social live streaming services (such as Periscope), and services for knowledge exchange (such as Twitter) (Linde & Stock, 2011, pp. 259ff.). In contrast to such media as newspapers, radio, or TV, in social media there is no formal information dissemination institution (as, say, The *New York Times*, CBS Radio, or NBC); thus, disintermediation happens. All social media are not immune from fake news (Zimmer, Scheibe, Stock, & Stock, 2019).

A *user* of Internet services acts as consumer (only receiving content), producer (producing and distributing content), and participant (liking or sharing content) on all kinds of online media (Zimmer, Scheibe, & Stock, 2018). In classical

communication science one speaks of the *audience* of media; nowadays, especially on social media, audience members are called "users." *Algorithms* are sets of rules defining sequences of operations; they can be implemented as computer programs in computational machinery. In this article, the term "algorithm" is only used in the context of computer programs running on "machines."

Filter bubbles and echo chambers are metaphorical expressions. For Pariser (2011), a *filter bubble* is a "unique universe of information for each of us." Pariser lists three characteristics of the relationship between users and filter bubbles, namely (1) one is alone in the bubble, (2) the bubble is invisible, and (3) the user never chose to enter the bubble. We will critically question Pariser's characteristics. For Dubois and Blank (2018, p. 3) a filter bubble means "algorithmic filtering which personalizes content presented on social media." Davies (2018, p. 637) defines filter bubbles as "socio-technical recursion," i.e. as an interplay between technologies (as, for instance, search engines or social media services) and the behavior of the users and their social relations.

An *echo chamber* describes "a situation where only certain ideas, information and beliefs are shared" (Dubois & Blank, 2018, p. 1). Echo chambers occur "when people with the same interests or views interact primarily with their group. They seek and share information that both conforms to the norms of their group and tends to reinforce existing beliefs" (Dubois & Blank, 2018, p. 3). Users in echo chambers are on a media or content "diet" (Case & Given, 2018, p. 116) or in "ideological isolation" (Flaxman, Goel, & Rao, 2016, p. 313) concerning a certain topic. Such isolation may result from *selective exposure of information* (Hyman & Sheatsley, 1947; Liao & Fu, 2013; Spohr, 2017) and a *confirmation bias* (Vydiswaran, Zhai, Roth, & Pirolli, 2015; Murungi, Yates, Purao, Yu, & Zhan, 2019). There are different manifestations of selective information exposure; its strongest form is "that people prefer exposure to communications that agree with their pre-existing opinions" (Sears & Freedman, 1967, p. 197). A special kind of selective exposure of information is "partisan selective exposure," which is related to political affiliations and not—as general selective exposure—based on ideologies or opinions (Kearney, 2019).

Both basic concepts are closely related; however, an *echo chamber* is more related to human information behavior and a *filter bubble* is more associated with algorithmic information filtering and results' presentation in online services.

Social media documents are skipping the *intermediation process*; indeed, "social media enabled a direct path from producers to consumers of contents, i.e., disintermediation, changing the ways users get informed, debate, and shape their opinions" (Bessi et al., 2015, p. 1). Prima facie, this sounds great. However, if we take a look at the other side of the coin, "confusion about causation may encourage speculations, rumors, and mistrust" (Bessi et al., 2015, p. 1). The disappearance of intermediation has not only "fostered a space for direct meetings in a sort of online Habermasian public sphere" (Törnberg, 2018, p. 17), but has also fostered misuse of social media through the publication of fake news by biased users. Habermas himself was always pessimistic about social media (Linde & Stock, 2011, p. 275), as for him weblogs play "a parasitical role of online communication" (Habermas, 2006, p. 423). The disappearance of intermediation also supports the parasitical roles of fake news in social media.

3. RESEARCH MODEL

The different estimations on the causes of fake news dissemination in social media directly lead to our central research question (RQ): Are echo chambers and filter bubbles of fake news man-made or produced by algorithms? To be more precise:

- RQ1: Is the dissemination of fake news supported by machines through the automatic construction of filter bubbles, and if yes, how do such algorithms work?
- RQ2: Are echo chambers of fake news man-made, and if yes, what are the information behavior patterns of those individuals reacting to fake news?

In our research model (Fig. 1), RQ1 is located on the left-hand side and RQ2 on the right hand side. We start searching for false propositions, i.e. fake news, and their dissemination via social media channels. First, we are going to describe processes leading to filter bubbles. A user will be informed of the existence of the false propositions via the push service of the social media platform. The selection of the documents which are shown to the user is controlled by the service's algorithms, which in turn are fed by the user's information behavior patterns and their behavior on the specific service (e.g., forming friendships, giving likes, etc.). It is possible that the interaction between the algorithms and the former user behavior clips only certain aspects of information content while neglecting all other content, thus forming a filter bubble. On Facebook, it is difficult to handle a bypass of the systems' algorithms. However, on other social media services, for instance, weblogs, there is a direct push of (fake) news to users. Following, we direct our attention to echo chambers. The same user can comment on the false propositions or reply to comments about such fake news.

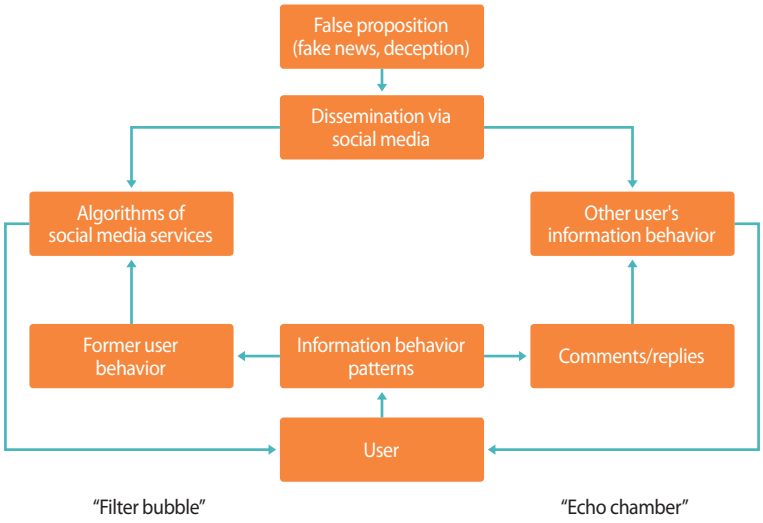


Fig. 1. Our research model: Filter bubble and echo chamber in social media.

His or her cognitive information behavior patterns may lead to different reactions such as confirmation, denial, moral outrage, and satire. In combination with other users' information behavior (replying to the user's comments or replies, liking them, sharing them, and so on) echo chambers of like-minded users may appear.

As there are two different research questions, this study applies different methods answering them. RQ1 will be evaluated by analyzing the sorting and presentation algorithms of social media by the example of Facebook. For RQ2 the authors performed empirical case study research applying content analysis of comments and replies on fake news distributed via social media channels. The channels disseminating the fake news were a weblog (*The Political Insider*) and two subreddits of the news aggregator Reddit, namely *r/The_Donald* and *r/worldpolitics*. We choose the blog from *The Political Insider* as it published the fake story on our case ("Hillary Clinton sold weapons to the Islamic State") for the first time; the subreddit *r/The_Donald* is clearly addressed to supporters of Donald Trump, while *r/worldpolitics* is a more liberal subreddit. As a result of this selection we were able to analyze comments from different ideological orientations.

How is our article structured? In the next paragraph, we define our basic terms. As fake news disseminate false propositions, it is necessary to discuss the concept of "truth" in relation to knowledge and information as well as to mediated contexts. In order to analyze and answer RQ1 this paper introduces relevance, pertinence, and ranking algorithms and describes Facebook's sorting algorithm in detail. To work on RQ2, we empirically studied patterns of cognitive processes of human information behavior in response to fake news. A case study provides us with empirical data of user comments and replies. Then, we describe the applied methods (case study research and content analysis), the empirical findings, and the data analysis. The final paragraph summarizes the main results, confesses limitations, and gives an outlook on further research.

4. KNOWLEDGE, INFORMATION, AND TRUTH

If we want to distinguish between fake (misinformation and disinformation) and non-fake (knowledge) we should know what knowledge, information, and truth are. The corresponding discipline is philosophy, more precisely epistemology. What follows is an excursus on the philosophical foundations of truth. The aim of this paragraph is to show that the definition of truth and the assignment of truth values to empirical statements are anything but easy.

Only a proposition is able to be true or false. In epistemology, one kind of knowledge ("knowing that" in contrast to "knowing how") is based on true propositions. Chisholm (1977, p. 138) defines knowledge:

h is known by *S* =df *h* is accepted by *S*; *h* is true; and *h* is nondefectively evident for *S*,

where *h* is a proposition and *S* a subject; =*df* means "equals by definition." Hence, Chisholm demands that the subject S accepts the proposition *h* (as true), which is in fact the case

(objectively speaking) and that this is so not merely through a happy coincidence, but precisely "nondefectively evident." Only if all three determinants (acceptance, truth, and evidence) are present, knowledge can be seen as well and truly established. In the absence of one of these aspects, such a statement can still be communicated—as information—but it would be an error (when truth and evidence are absent), a supposition (if acceptance and evidence are given, but the truth value is undecided) or a lie, fake, or deception (when none of the three aspects apply).

Knowledge cannot be transmitted as such; it is in need of a sender, data to be transmitted, a channel, and a receiver. Information dynamically sets knowledge "into motion." Knowledge always has a truth claim. Is this also the case for information, if information is what sets this knowledge in motion? Is there something like true or false information (Stock & Stock, 2013, p. 39)? Apart from knowledge, there are further, related forms of dealing with objects. If beliefs, conjectures, or fakes are put into motion, are they not information? "Information is not responsible for truth value," Kuhlen (1995, p. 41) points out. Buckland (1991, p. 50) remarks, "we are unable to say confidently of anything that it could not be information." Maybe the proposition which is transmitted by information is true or "contingently truthful" (Floridi, 2005); and many information scientists "will generally ignore any distinction between truth or falsity of information" (Case & Given, 2018, p. 67). The task of checking the truth value of the knowledge, rather, must be delegated to the receiving subject S. She or he then decides whether the information retrieved represents knowledge, conjecture, or untruth. Therefore, it is terminologically very problematic to speak of "true/false *information*," as only *propositions* are truth bearers.

Propositions, linguistically presented by declarative sentences, can be true or false. Here, one basic philosophical question arises. Even Pontius Pilate once famously asked "What is truth?" to which Jesus responded—with silence. Truth is a relation between a proposition and a reference object. There are different truth theories working with different reference objects, namely reality, praxis, other propositions in the same system, acceptance inside a community, and, finally, a person's internal state.

The classical approach to analyze truth is the *correspondence theory* (David, 1994) theorizing the relation between a proposition and a concrete fact in space and time. Although there are similar definitions of correspondence already in Aristotle's work, the canonical form of this truth theory originates from the early twentieth century. Bertrand Russell states, "(t)hus a belief is true when there is a corresponding fact and is false when there is no corresponding fact" (Russell, 1971, p. 129). A person, who will make true propositions on a certain state of affairs in reality, must perceive (watch, hear, etc.) this part of reality personally, in real-time, and on site. In our context of journalism and social media, the person reporting on a state of affairs makes a true proposition ("true" for his self-consciousness) when he luckily is in the right spot at the right time. In times of social media, the term "journalist" includes professional investigative journalism as well as citizen journalists reporting via channels like Facebook, Reddit, Twitter, or Periscope. For the audience of those journalists, there is no chance to verify or to falsify the correspondence between the read or heard proposition in the newspaper, the tweet, or the TV broadcast, and the part of reality, since they simply were not there. This is the reason why the correspondence theory of truth only plays a minor role, if any, in the context of fake or alternative news (Muñoz-Torres, 2012).

Accordance with objective reality and personal awareness is the key factor of the *theory of reflection*. Whether the human mind contains truth is not a question of theory, but of praxis. In praxis (working, any decision procedure), humans have to prove the truth of their thinking in their practical behavior (Pawlow, 1973). A sentence is true if its proposition works in practice. The problem with the theory of reflection is that it is impossible to consider all facts because they are always a product of selection. A problem of the media is that it sometimes takes a while to gather all facts to accurately use them in practice. By the time the facts were gathered the media momentum has passed.

The *coherence theory* of truth declares that one statement corresponds with another statement, or with the maximal coherent sum of opinions and accepted clauses of statements (Neurath, 1931). There cannot be an opposite statement within an already accepted system of statements. If the statement can be integrated, it is true, otherwise it is false. However, instead of rejecting the new statement, it is possible to change the whole system of statements to integrate the latest one into the system. The statements need to be logically derivable from each other.

The definition of the *consensus theory* of truth states that truth is what is agreed upon by all people in a group. First, the speakers need to be clear about what they are saying to ensure everyone understands what they mean, they insinuate each other's truthfulness, and their words are accurate. A discourse needs to determine if the claim of the speaker is indeed to be accepted. Everyone needs to have the same level of influence to rule or to oppose (Habermas, 1972). Relying only on the consensus theory of truth is difficult and does not necessarily lead to the truth in the sense of the correspondence theory.

Brentano (1930) describes the *evidence theory* of truth, "When I have evidence, I cannot err." A judgement is true if

it expresses a simple quality of experience. Brentano adheres to the traditional view that there are two different ways for a judgement to be evident; either it is immediately, or it is evident insofar as it is inferable from evident judgements by applications of evident rules. But, evidence is a primitive notion; it cannot be defined, it is only experienceable, and thus, found in oneself.

The philosophical truth theories illustrate that truth or lies are in the eye of the beholder (evidence theory), the praxis (theory of reflection), the community (consensus theory), or in the system of accepted propositions (coherence theory). As the correspondence theory of truth is not applicable in the environments of journalism and social media, we have big problems in stating what exactly is true and what is not. If we do not know what *the* truth is, we also cannot know exactly what "fake news" is. It is the individual person who decides, based on a (probably unknown) truth theory, what is considered as truth, as lies, as "true news," and as "fake news." By the way, attempts of automatic semantic deception detection (e.g., Conroy, Rubin, & Chen, 2015) are faced with the same problems, especially when they rely on the coherence or the consensus theory of truth.

5. FAKE NEWS DISSEMINATION THROUGH ALGORITHMIC FILTER BUBBLES (RQ1)

The concept of *relevance* is one of the basic concepts of information science (Saracevic, 1975). Users expect an information system to contain relevant knowledge, and many information retrieval systems, including Internet search engines and social media services, arrange their results via relevance ranking algorithms. In information science, researchers distinguish between objective and subjective information needs. Correspondingly to these concepts, we speak of relevance (for the former) and pertinence (for the latter), respectively.

Since relevance always aims at user-independent, objective observations, we can establish a definition: A document, for instance, a website, a blog post, a post on Facebook or Reddit, or a microblog on Twitter (or, to speak more precisely, the knowledge contained therein) is *relevant* for the satisfaction of an objective (i.e. subject-independent) information need.

A research result can only be *pertinent* if the user has the ability to register and comprehend the knowledge in question according to his or her cognitive model. Soergel (1994, p. 590) provides the following definition: "Pertinence is a relationship between an entity and a topic, question, function, or task with respect to a person (or system) with a given purpose. An entity is pertinent if it is topically relevant and if it is appropriate for the person, that is, if the person can understand the document and apply the information gained." Pertinence ranking presupposes that the information system in question is able to identify the concrete user who works with the system; it is always subject-dependent personalized ranking (Stock & Stock, 2013, pp. 361ff.).

We describe only one paradigmatic example of ranking in social media, namely the algorithms of Facebook as the most common social media platform. Facebook's sorting of posts is a pertinence ranking algorithm; it works with the three factors affinity, weighting, and timeliness. According to these three aspects, a user will see posts on her or his Facebook page with the posts sorted in descending order of their retrieval status values (Zuckerberg et al., 2006). Affinity is concerned with the user's previous interactions on the posting pages, whereas different interactions are weighted variously. If a user *X* frequently views another user's (say, user *A*) posts, likes them, comments on them, or shares them, *A*'s future posts—depending on their weights (resulting from the numbers of likes, shares, and comments)—get a higher weight for user *X*. Facebook also considers the position of the creator of the post (is this user often viewed, annotated, etc.?) and the nature of the post (text, image, or video). The timeliness states that a contribution becomes more important the newer it is. However, other factors play a role, and the algorithm is constantly being adapted. For example, an already viewed ranked list is not displayed a second time in exactly the same order (i.e., the criteria for the sorting are each slightly modified) in order to make the lists more interesting. Also, posts from people (as opposed to those from companies) are weighted higher, and the spatial proximity between the receiver and the sender of the post plays an important role. In particular, the affinity causes a user to see the one source at the top of his or her list, which he or she has often viewed in previous sessions.

Ranking on Facebook is always personalized and based on the user's common interests, her or his information behavior on the service, and her or his Facebook friends (Tseng, 2015; Bakshy, Messing, & Adamic, 2015). The more a user repeatedly clicks on the posts of the same people, the more the selection of posts stabilizes, which always appear at the ranking's top positions. Thus, in a short time—with high activity on Facebook—an information diet may occur that presents users only those posts on top of their pages, whose creators they prefer. So it can be assumed that such personalized content representation leads to "partial information blindness (i.e., filter bubbles)" (Haim, Graefe, & Brosius, 2018, p. 330).

It depends on the user to form a "friendship" on Facebook, and it is on the user to often select certain friends' posts

for reading, liking, sharing, and commenting. Facebook's pertinence ranking algorithm indeed may amplify existing behavioral patterns of the users into filter bubbles and then into echo chambers, whereby the information behavior of the users plays the important primary role. In contrast to the assumptions of Pariser (2011) on filter bubbles, (1) no one is alone in the bubble when the bubble leads to echo chambers (where other users are by definition); (2) the bubble is visible to certain users insofar as they figured out Facebook's ranking methods; for other, rather uncritical users, the bubble is indeed invisible; (3) the users' behavior feeds the pertinence ranking algorithms; therefore, the users (consciously or unintentionally) cooperate with the service entering the bubble through their own information behavior.

Here we arrive at a first partial result and are able to answer RQ1: Algorithms by themselves do not produce filter bubbles or subsequently echo chambers, they only consolidate the users' information behavior patterns. Concerning the reception of fake news, it is not possible to argue that they are solely distributed by "bad algorithms," but by the active collaboration of the individual users. Also, Del Vicario et al. (2016, p. 554f.), for instance, found out that "content-selective exposure is the primary driver of content diffusion and generates the formation of homogeneous clusters, i.e., 'echo chambers.'" DiFranzo and Gloria-Garcia (2017, p. 33f.) arrive at a similar result: "The related filter-bubble effect is due to the user's network and past engagement behavior (such as clicking only on certain news stories), that is, it is not the fault of the news-feed algorithm but the choices of users themselves." There are results concerning fake news and the algorithms of Facebook: "While this criticism has focused on the 'filter bubbles' created by the site's personalisation algorithms, our research indicates that users' own actions also play a key role in how the site operates as a forum for debate" (Seargeant & Tagg, 2019, p. 41). Although algorithms are able to amplify human information behavior patterns, obviously, the users play the leading role concerning construction and maintenance of those bubbles of (fake) news. Indeed, there are filter bubbles; however, they are fed by users' information behavior and—more important—they are escapable (Davies, 2018).

6. FAKE NEWS DISSEMINATION THROUGH MAN-MADE ECHO CHAMBERS (RQ2)

6.1. Our Approach

When we want to analyze echo chambers of fake news and also believing as well as mistrusting such false propositions by individual persons, we have to study their cognitive processes in detail. In our research study, we apply case study research and content analysis. As we want to investigate which concrete cognitive information behavior patterns concerning fake news exist, we start our endeavors with the help of concrete cases. Case study researchers "examine each case expecting to uncover new and unusual interactions, events, explanations, interpretations, and cause-and-effect connections" (Hays, 2004, p. 218f.). Our case includes a (probably fake) post and comments as well as replies to it. It is a story on Hillary Clinton selling weapons to the Islamic State. With the help of this singular case study (Flyvbjerg, 2006) we try to find cognitive patterns and to understand users' information behavior at the time shortly after the publication of fake news.

To analyze the cognitive patterns of the commenting users, we look upon the results of the cognitive processes, i.e. the texts (as we are not able to measure the human cognitive patterns directly) and apply quantitative and qualitative content analysis (Krippendorff, 2018) of posts in social media. In quantitative content analysis, the occurrence of the categories in the coding units is counted and, if necessary, further processed statistically; the qualitative content analysis turns to the statements within the categories, namely the "manifest content" (Berelson, 1952) and the "deeper meaning" (such as subjective senses), as well as formal textual characteristics such as style analysis (Mayring & Fenzl, 2019). In order to create the appropriate categories for the content analysis, we applied both (1) inductive (or conventional) as well as (2) deductive (or directed) measures (Elo & Kyngas, 2008; Hsieh & Shannon, 2005). By (1) applying the conventional approach with a first and preliminary analysis of comments concerning our case, we defined the first codes; and we (2) arrived at codes while studying relevant published literature. The coding unit was the single comment or the single reply. Every coding unit was coded with only one (the best fitting) category. The coding process was led by a short code book and conducted by two of the article's authors in August 2018, whereas all steps were performed intellectually. In a first round, the coders worked independently (resulting in Krippendorff's alpha > 0.8, signaling the appropriateness of the code book and the coders' work); in a second round, the (few) disagreements were discussed and solved (Mayring & Fenzl, 2019, p. 637). In the end, there was an intercoder consistency of 100%, i.e., Krippendorff's alpha was 1.

Our approach is similar to research in microhistory describing posts and comments on social networking services in order to find information on historically relevant—especially local—events and developments (Stock, 2016, 2017). Similar

to our approach, Walter, Brüggemann, and Engesser (2018) studied user comments in echo chambers concerning the topic of climate change. Gilbert, Bergstrom, and Karahalois (2009) defined agreement as manifestation of an echo chamber. They found that about 39% of all comments agree with the blog author, 11% disagree, and half of all commentators react in other ways. Murungi et al. (2019, pp. 5192f.) found that significant amounts of comments on a concrete political situation (Roy Moore's candidacy for the U.S. Senate in Alabama in 2017) were non-argumentative.

For our case study, we consulted a weblog (*The Political Insider*, a right-wing oriented web site; August 2016) (N=43) and *Reddit* as the current most popular news aggregator (Zimmer, Akyürek et al., 2018). To be more precise, we analyzed Reddit's subreddits *r/The_Donald* (a forum "for Trump supporters only"; September 2016) (N=177) and *r/worldpolitics* (a "free speech political subreddit"; September 2016) (N=246). We checked all comments and all replies to the comments manually. All in all, we analyzed 466 documents. Studying literature and empirical material, we found different patterns of information behavior in response to fake news and applied them as codes for our content analysis:

- Confirmation: broad agreement with post, attempt of verification
- Denial: broad disagreement with post, attempt of falsification
- Moral outrage: questioning the posts, comments and replies from a moral point of view
- New rumor: creation of a new probably false proposition
- Satire: satirical, ironic, or sarcastic text
- Off-topic: non-argumentative, ignoring the discussion, arguing on other topics, broad generalization
- Insult: defamation of other people or groups
- "Meta" comment/reply: discussing the style of another post, offense against a commentator

Additionally, we evaluated the topic-specific orientation (positive, negative, and neutral) for all texts. *Positive* means an articulated or implicated agreement with the original post. If a comment, for instance, argues, "Clinton should be arrested" in response to the post "Hillary Clinton sold weapons to ISIS," it is counted as positive. *Neutral* means that there is no relation to the concrete topic of the triggering post, e.g., "Obama is born in Kenya" as a comment on "Clinton sold weapons." All other texts were coded as *negative*, e.g., "What's there to say? It's just a vague, unfounded accusation." We aggregated all generations of replies (replies to a comment, replies to a reply) into the code "reply."

6.2. Results

Tables 1-3 exhibits our descriptive results for the three selected sources, namely *The Political Insider*, *r/The_Donald*, and *r/worldpolitics*. Concerning our case study, most comments on *The Political Insider* are confirmations of the (false) proposition; likewise, the comments' orientation is predominantly positive (Table 1). In both analyzed subreddits most comments (about 40% to 50%) and even more replies (about 70% to 80%) are non-argumentative or off-topic (Tables 2 and 3). In the subreddit *r/The_Donald* we found about 40% agreement with the fake proposition for the comments; however, only 8% existed for the replies.

About half of the comments in *r/The_Donald* express a neutral orientation, and the other half a positive one; while most of the replies were neutral. Most comments and more than 80% of the replies in *r/worldpolitics* are off-topic and express no orientation concerning the given topic (i.e., the triggering post). The authors of *r/worldpolitics* are more critical than those of *r/The_Donald* as about 30% of all comments were classified as denial (in contrast to 0% in *r/The_Donald*).

The dominating cognitive patterns are *non-argumentative* or arguments being *off-topic*. The very first comment on *r/worldpolitics* was "time to put up or shut up," which diverse authors regarded as an invitation to speculate on different political topics with loose or no relationship to the content of the post. We can find rather senseless texts as, e.g., "LOL who knew," "Holy shit!!", or "Trump was right all along" (all from *r/The_Donald*). However, most of the off-topic comments and replies pursue a similar tendency, most notably attacking Obama and praising Trump in *r/The_Donald* or discussing the DNC (Democratic National Committee) in *r/worldpolitics*.

Confirmations of the fake news are frequent in *The Political Insider* and *r/The_Donald*, but not in *r/worldpolitics*. Here are some examples: "Done, done, DONE! Round up his people"—"Traitors are hanged from the highest tree!"—"His eyes were always cold to me … soulless. It is no surprise that Obama would be the founder of ISIS, really." Confirmations culminate in death threats: "Put him [i.e., Obama] to death. Period. Let the left cry. They will never agree that they are wrong, that he was a criminal. It doesn't matter. He is a traitor to this country, and if these allegations are true, he needs to be appropriately punished" (all from *r/The_Donald*).

Sometimes, commentators are dissatisfied with the discussion and argue from a *meta position* as "I'm really not interested in engaging in a totally off-topic argument with you"; "What? Seriously you believe this?"; or "Why have you sent me an article about how George Bush, the Republican president, may have rigged the 2004 election as evidence that Hillary Clinton, the

Table 1. Users' cognitive patterns in reactions to fake news: *The Political Insider*

Cognitive pattern	Comments	Replies
Confirmation	33.3%	23.1%
Denial	3.3%	-
Moral outrage	3.3%	-
New rumor	13.3%	15.4%
Satire	-	-
Off-topic	26.6%	61.5%
Insult	20.6%	-
"Meta"	-	-
Positive orientation	73.3%	46.2%
Negative orientation	3.3%	-
Neutral orientation	23.3%	53.8%
N	30	13

Post: "Wikileaks CONFIRMS Hillary Sold Weapons to ISIS… Then Drops Another BOMBSHELL! Breaking News."

Table 2. Users' cognitive patterns in reactions to fake news: *r/The_Donald*

Cognitive pattern	Comments	Replies
Confirmation	40.8%	7.9%
Denial	-	4.0%
Moral outrage	-	-
New rumor	5.3%	5.0%
Satire	1.3%	2.0%
Off-topic	47.4%	78.2%
Insult	5.3%	3.0%
"Meta"	-	-
Positive orientation	48.7%	11.9%
Negative orientation	-	5.0%
Neutral orientation	51.3%	83.2%
N	76	101

Post: "Breaking Assange: Obama & Clinton not only supplied ISIS with a billion dollars worth of weapons annually, they paid these mercenaries salaries! Obama employed ISIS… let it sink in. Obama was the real leader of ISIS!"

Table 3. Users' cognitive patterns in reactions to fake news: *r/worldpolitics*

Cognitive pattern	Comments	Replies
Confirmation	12.5%	9.1%
Denial	29.2%	6.1%
Moral outrage	-	1.0%
New rumor	2.1%	0.5%
Satire	4.2%	0.5%
Off-topic	43.8%	72.2%
Insult	2.1%	0.5%
"Meta"	6.3%	10.1%
Positive orientation	14.6%	9.6%
Negative orientation	31.3%	6.6%
Neutral orientation	54.2%	83.8%
N	48	198

Post: "Julian Assange: '1,700 emails' proves Hillary Clinton sold weapons to ISIS in Syria."

Democratic candidate, has rigged the upcoming election?" (all from *r/worldpolitics*).

Some (however few) comments are *insults*, as, for instance, "Yet more proof that the people at the very top are, for all practical purposes, gangsters" (*r/worldpolitics*); "Obama is a piece of shit Globalist muslim"; or "Aw, come on. Whadya expect from a f**kin' Kenyan 'born' in Hawaii, raised in Indonesia, programmed and sponsored by the Saudi Manchurian School for Gifted Leftists?" (both from *r/The_Donald*).

Here, a further cognitive pattern comes into play: the construction of a *new rumor*, for example: "The Hawaiian birth certificate (of Obama, a/n) was proven to be a forgery"; "Obama's entire life is pure fiction, a 100% CIA creation"; "Hillary is the Mother of ISIS"; "They (Obama and Clinton, a/n) wanted this war in Syria, they wanted the refugee influx"; or, "It will take a while before people admit that Obama and Michelle and the supported 'daughters' were all fake"; "Malia's and Sasha's biological parents have always been nearby while the girls provided a fictional family for Barack and Michelle" (all from *r/The_Donald*).

Some comments and replies consist of satirical, ironic, or sarcastic text, as for instance: "Of course, president Hussein was the head of Isis. He's a muzlim [sic]" (*r/The_Donald*); "Is that really how your brain works? Or you just playin'?"; "Someone feel like pointing to some of those emails? Julian? Anybody? Like most Americans, I am too stupid and lazy to spend four years reading emails"; "This news article is great, and absolutely 100% real. I can't wait to see this actually real story break worldwide, because Hillary absolutely sold weapons to ISIS in Syria, and this is not at all a conspiracy theory!" (all from *r/worldpolitics*). Sometimes it is problematic to identify irony; however, considering the context the pattern becomes visible.

In the subreddit *r/worldpolitics* (but with next to nothing in *The Political Insider* and *r/The_Donald*) we found critical *denials* of the fake news as, for instance, "get suspicious when it's only niche websites reporting stuff like this. If there were real evidence, every conservative site would make a front page"; or, "1700 mails about Libya proof that Hillary sold weapons to Isis in Syria? I don't mean to comment on the allegations but I hate it when headlines are clearly bullshit."

A rather uncommon pattern in this case study is *moral outrage*, a kind of meta-comment from a moral point of view, for instance: "All of you are blaming Hillary and President Obama. They have to get approval from Congress to do this stuff" (*The Political Insider*); or, "What's there to say? It's just a vague, unfounded accusation" (*r/worldpolitics*).

There are different distributions of cognitive patterns regarding the level of discussion, i.e. between the first generation

of texts (comments on the triggering fake news) and the next generations (replies to the comments and replies to other replies). There are much more non-argumentative and off-topic replies than off-topic comments (*The Political Insider*: 62% versus 27%; *r/The_Donald*: 78% versus 47%; *r/worldpolitics*: 72% versus 44%). And there are less confirmative replies than confirmative comments (*The Political Insider*: 23% versus 33%; *r/The_Donald*: 8% versus 41%; *r/worldpolitics*: 9% versus 13%). Additionally, the users' information behavior is drifting from positive or negative orientation at the comments' level to an enhanced neutral orientation at the replies' level.

6.3. Are There Indeed Echo Chambers?

What can we learn from our case study? Do users indeed live inside an echo chamber? The answer depends on the concrete operationalization of the "echo chamber." If we narrowly define this concept as a community with high confirmation rates (in our case: for fake news) in combination with high degrees of positive topic-specific orientation (and further with the creation of new rumors with the same direction as the original fake), there are indeed hints for the existence of such communities. A third of the commentators of *The Political Insider* and about two-fifths of the commenting audience of *r/The_Donald* seem to argue inside their echo chambers. However, we can define "echo chamber" more broadly. As we know from the texts, off-topic comments and most of the neutral-orientation texts argue in the same direction as the entire community; therefore the filter bubble may include most of these comments and replies. The content of the specific (false) proposition is entirely clear and taken for granted, so users lose the specific thread (from the triggering post); however, they do not lose the (ideological or political) direction. In the sense of this broad definition, depending on the source, up to about 90% of comments (sum of confirmations and off-topic comments) in *r/The_Donald*, about 60% in *The Political Insider*, and about 55% in *r/worldpolitics* exhibit hints towards the existence of echo chambers in those social media channels. In contrast to Bruns (2019) we found that the problems concerning filter bubbles and echo chambers are not overstated, but basic facts in our contemporary online world.

7. CONCLUSION

As the correspondence theory of *truth* is not applicable in mediated contexts, there remain truth theories which heavily depend on the community (consensus theory) and on the coherence of propositions (coherence theory), but do not point to *the* truth. This annoying fact does not make research on fake news easy.

Algorithms (and their mechanisms to form filter bubbles) applied in social media themselves do not form communities purely on their own as they amplify users' information behavior. The crucial element of fake news and their pathways into social media is mainly the individual *users*, their cognitive patterns, and their surrounding echo chamber (Zimmer, 2019).

Reading (fake) news and eventually drafting a comment or a reply may be the result of users' selective exposure to information (Frey, 1986; Sears & Freedman, 1967) leading to preferring news (including fake news) fitting their pre-existing opinions. If users take the (false) proposition as given, discuss it uncritically, ignore other opinions, or argue further off-topic (however, always in the same direction), an echo chamber can be formed and stabilized. In contrast to some empirical findings on echo chambers (Fischer et al., 2011; Garrett, 2009; Nelson & Webster, 2017) we found clear hints for the existence of such communities. Depending on the concrete operationalization of the "echo chamber," about one third to two-fifths (a narrow definition) and more than half of all analyzed comments and replies (a broad definition) can be located inside an echo chamber of fake news. Explicitly expressed confirmation depends on the stage of discussion. In the first stage (comments), confirmative texts are more frequent than in further stages (replies).

Confirmative information behavior on fake news goes hand in hand with the consensus and the coherence theory of truth. The (in the sense of the correspondence theory of truth basically false) proposition will be accepted "by normative social influence or by the coherence with the system of beliefs of the individual" (Bessi et al., 2015, p. 2). This behavior leads directly to a confirmation bias. Our results are partly in line with the theory of selective exposure of information.

However, it is not possible to explain *all* information behavior following fake news with the theory of selective exposure, but with a variety of further individual cognitive patterns. We were able to identify cognitive patterns clearly outside of echo chambers as denial, moral outrage, and satire—all in all patterns of critical information behavior.

This study has (as every scientific endeavor) *limitations*. In the empirical part of the study, we analyzed comments and replies to comments on social media. The publication of a comment or a reply on an online medium follows a decision-making process (should I indeed write a comment or a reply?). With our method, we are only able to gather data on individuals who have written such texts; all others remain unconsidered. We did not talk to the commenting and replying individuals. Therefore, we

were not able to ask for intellectual backgrounds, motivations, and demographic details of the commentators.

In this article, we report about one case study only, so the extent of the empirical data is rather limited. Although we collected and intellectually coded some hundreds of texts, this is like a drop in the bucket when faced with millions of posts, comments, and replies on social media. A serious methodological problem (not only ours, but of all research relying on data from the Internet) is the availability of complete data sets on, for instance, a fake news story and all the comments and replies on the fake news, as users and website administrators often delete discriminating posts, comments, or replies. We indeed found hints for deleted posts, comments, or replies on *The Political Insider* as well as on *Reddit*. In lucky cases (as in our study: the post and the comments of *The Political Insider*), one will find some deleted data on web archives.

Here are some recommendations for future research. As we only analyzed texts on fake news in order to find cognitive reaction patterns, research should also study in analogous ways reactions to true propositions. Are there the same cognitive patterns? People do not only live in the online world. Of course, their lives in the physical world are influenced by family members, friends, colleagues, and other people. As there are empirical hints on the geographic embedding of online echo chambers (Bastos et al., 2018), it would be very helpful to analyze offline echo chambers and the interplay between online and offline echo chambers as well. We distinguished between comments and replies and found different cognitive patterns of the respective authors. Are there indeed different cognitive patterns while writing posts, formulating comments, and phrasing replies to the comments? How can we explain those differences?

What is new in this paper? As algorithms (as, for instance, Facebook's ranking algorithm) only amplify users' information behavior, it is on the individuals themselves to accept or to deny fake news uncritically, to try to verify or to falsify them, to ignore them, to argue off-topic, to write satire, or to insult other users. If filter bubbles are made by algorithms and echo chambers by users, the echo chambers influence the filter bubbles; however, filter bubbles strengthen existing echo chambers as well. There are different cognitive patterns of the individual users leading to different reactions to fake news. Living in echo chambers (namely the uncritical accepting of the news due to the users' pre-existing opinions shared within a group or compared with a set of propositions) indeed is a typical, but not the only cognitive pattern.

Therefore, a "critical user" seems to be the decisive factor in identifying and preventing fake news. Our analysis at the beginning of this paper has shown that there is no satisfying answer to what can be considered *the* truth in media. In the end—and this is in line with Chisholm's (1977) definition of knowledge—it is just a critical user who compares sources and validates the timeliness and evidence of a contribution before believing, denying, or ignoring it and then deciding whether it is true or false. So, finally, it is on the individual user's critical literacy, information literacy, digital literacy, and media literacy in order "to help cultivate more critical consumers of media" (Mihailidis & Viotty, 2017, p. 441) and, additionally, on libraries and information professionals to instruct their users "in the fight against fake news" (Batchelor, 2017, p. 143) and to "become more critical consumers of information products and services" (Connaway, Julien, Seadle, & Kasprak, 2017, p. 554). Libraries, next to schools (Gust von Loh & Stock, 2013), are faced with the task to educate and instruct people to become critical users.

REFERENCES

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211-236.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Political science: Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.

Bastos, M., Mercea, D., & Baronchelli, A. (2018). The geographic embedding of online echo chambers: Evidence from the Brexit campaign. *PLoS One*, 13(11), e0206841.

Batchelor, O. (2017). Getting out the truth: The role of libraries in the fight against fake news. *Reference Services Review*, 45(2), 143-148.

Berelson, B. (1952). *Content analysis in communications research*. Glencoe: Free Press.

Bessi, A., Zollo, F., Del Vicario, M., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Trend of narratives in the age of misinformation. *PLoS ONE*, 10(8), e0134641.

Brentano, F. (1930). *Wahrheit und Evidenz*. Leipzig: Meiner.

Bruns, A. (2017). Echo chamber? What echo chamber? Reviewing the evidence. In *6th Biennal Future of Journalism Conference (FOJ17)*, 14-15 September, Cardiff, UK. Retrieved Apr 18, 2019 from https://eprints.qut.edu.au/113937/.

Bruns, A. (2019). *Are filter bubbles real?* Cambridge: Polity Press.

Buckland, M. K. (1991). *Information and information systems*. New York: Praeger.

Case, D. O., & Given, L.M. (2018). *Looking for information: A survey of research on information seeking, needs, and behavior* (4th ed.). Bingley: Emerald.

Chisholm, R. M. (1977). *Theory of knowledge*. Englewood Cliffs: Prentice-Hall.

Connaway, L. S., Julien, H., Seadle, M., & Kasprak, A. (2017). Digital literacy in the era of fake news: Key roles for information professionals. *Proceedings of the Association for Information Science and Technology*, 54(1), 554-555.

Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1), 1-4.

David, M. (1994). *Correspondence and disquotation. An essay on the nature of truth*. Oxford: Oxford University Press.

Davies, H. C. (2018). Redefining filter bubbles as (escapable) socio-technical recursion. *Sociological Research Online*, 23(3), 637-654.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America*, 113(3), 554-559.

DiFranzo, D., & Gloria-Garcia, K. (2017). Filter bubbles and fake news. *XRDS: Crossroads, ACM Magazine for Students*, 23(3), 32-35.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729-745.

Elo, S., & Kyngas, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107-115.

Fischer, P., Lea, S., Kastenmuller, A., Greitemeyer, T., Fischer, J., & Frey, D. (2011). The process of selective exposure: Why confirmatory information search weakens over time. *Organizational Behavior and Human Decision Making*, 114(1), 37-48.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298-320.

Floridi, L. (2005). Is information meaningful data? *Philosophy and Phenomenological Research*, 70(2), 351-370.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219-245.

Frey, D. (1986). Recent research on selective exposure to information. *Advances in Experimental Social Psychology*, 19, 41-80.

Garrett, R. K. (2009). Echo chambers online? Politically motivated selective exposure among Internet new users. *Journal of Computer-Mediated Communication*, 14(2), 265-285.

Gilbert, E., Bergstrom, T., & Karahalios, K. (2009). Blogs are echo chambers: Blogs are echo chambers. In *Proceedings of the 42nd Hawaii International Conference on System Sciences* (pp. 1-10). Washington, DC: IEEE Computer Science.

Gust von Loh, S., & Stock, W. G. (2013). Informationskompetenz als Schulfach? In S. Gust von Loh & W. G. Stock (Eds.), *Informationskompetenz in der Schule. Ein informationswissenschaftlicher Ansatz* (pp. 1-20). Berlin: De Gruyter Saur.

Habermas, J. (1972). *Knowledge and human interests*. Boston: Beacon Press.

Habermas, J. (2006). Political communication in media society. Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. *Communication Theory*, 16(4), 411-426.

Haim, M., Graefe, A., & Brosius, H. B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism*, 6(3), 330-343.

Hays, P. A. (2004). Case study research. In K. deMarrais & S. D. Lapan (Eds.), *Foundations for research* (pp. 217-234). Mahwah: Lawrence Erlbaum.

Hern, A. (2017, May 22). How social media filter bubbles and algorithms influence the election. *The Guardian*, Retrieved Apr 18, 2019 from https://www.theguardian.com/technology/2017/may/22/social-media-election-facebook-filter-bubbles.

Holmes, R. (2016, December 8). The problem isn't fake news, it's bad algorithms: Here's why. *Observer*. Retrieved Apr 18, 2019 from https://observer.com/2016/12/the-problem-isnt-fake-news-its-bad-algorithms-heres-why.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.

Hyman, H. H., & Sheatsley, P. B. (1947). Some reasons why information campaigns fail. *Public Opinion Quarterly*, 11(3), 412-423.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59-68.

Kearney, M. W. (2019). Analyzing change in network polarization. *New Media & Society*, 21(6), 1380-1402.

Kim, A., & Dennis, A. R. (2018). Says who? How news presentation format influences perceived believability and the engagement level of social media users. In

Proceedings of the 51st Hawaii International Conference on System Sciences (pp. 3955-3965). Washington, DC: IEEE Computer Science.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). Los Angeles: Sage.

Kuhlen, R. (1995). *Informationsmarkt. Chancen und Risiken der Kommerzialisierung von Wissen*. Konstanz Germany: UVK.

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the 'post-truth' era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353-369.

Liao, Q.V., & Fu, W.T. (2013). Beyond the filter bubble: Interactive effects of perceived threat and topic involvement on selective exposure to information. In *Proceedings of the SIGHCI Conference on Human Factors in Computing Systems* (pp. 2359-2368). New York: ACM.

Linde, F., & Stock, W. G. (2011). *Information markets: A strategic guide for the i-commerce*. Berlin: De Gruyter Saur.

Mans, S. (2016, November 13). *In the filter bubble: How algorithms customize our access to information* [Web blog post]. Retrieved Apr 18, 2019 from https://mastersofmedia.hum.uva.nl/blog/2016/11/13/in-the-filter-bubble/.

Mayring, P., & Fenzl, T. (2019). Qualitative inhaltsanalyse. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 633-648). Wiesbaden: Springer Fachmedien.

Mihailidis, P., & Viotty, S. (2017). Spreadable spectacle in digital culture: Civic expression, fake news, and the role of media literacy in 'post-fact' society. *American Behavioral Scientist*, 61(4), 441-454.

Muñoz-Torres, J. R. (2012). Truth and objectivity in journalism. Anatomy of an endless misunderstanding. *Journalism Studies*, 13(4), 566-582.

Murungi, D. M., Yates, D. J., Purao, S., Yu, Y. J., & Zhan, R. (2019). Factual or believable? Negotiating the boundaries of confirmation bias in online news stories. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 5186-5195). Honolulu: HICSS.

Nelson, J. L., & Webster, J. G. (2017). The myth of partisan selective exposure: A portrait of the online news audience. *Social Media + Society*, 3(3). doi:10.1177/2056305117729314.

Neurath, O. (1931). Soziologie im Physikalismus. *Erkenntnis*, 2(1), 393-431.

Oxford Dictionaries (2016). *Word of the year 2016 is post-truth*. Retrieved Apr 18, 2019 from https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2016.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London: Viking.

Pawlow, T. (1973). *Die Widerspiegelungstheorie*. Berlin: Deutscher Verlag der Wissenschaften.

Quattrociocchi, W. (2017). Inside the echo chamber. *Scientific American*, 316(4), 60-63.

Russell, B. (1971). *Problems in Philosophy*. Oxford: Oxford University Press.

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26(6), 321-343.

Seargeant, P., & Tagg, C. (2019). Social media and the future of open debate: A user-oriented approach to Facebook's filter bubble conundrum. *Discourse, Context and Media*, 27, 41-48.

Sears, D. O., & Freedman, J. L. (1967). Selective exposure to information: A critical review. *Public Opinion Quarterly*, 31(2), 194-213.

Self, W. (2016, November 28). Forget fake news on Facebook: The real filter bubble is you. *New Statesman America*. Retrieved Apr 18, 2019 from https://www.newstatesman.com/science-tech/social-media/2016/11/forget-fake-news-facebook-real-filter-bubble-you.

Shin, J., Jian, L., Driscoll, K., & Bar, F. (2018). The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior*, 83, 278-287.

Soergel, D. (1994). Indexing and the retrieval performance: The logical evidence. *Journal of the American Society for Information Science*, 45(8), 589-599.

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*, 34(3), 150-160.

Stock, M. (2016). Facebook: A source for microhistory? In K. Knautz & K.S. Baran (Eds.), *Facets of Facebook: Use and users* (pp. 210-240). Berlin: De Gruyter Saur.

Stock, M. (2017). HCI research and history: Special interests groups on Facebook as historical sources. In *HCI International 2017: Posters' Extended Abstracts. HCI 2017* (pp. 497-503). Cham: Springer.

Stock, W. G., & Stock, M. (2013). *Handbook of information science*. Berlin: De Gruyter Saur.

Tandoc Jr., E. C., Lim, Z. W., & Ling, R. (2018). Defining 'fake news.' A typology of scholarly definitions. *Digital Journalism*, 6(2), 137-153.

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS ONE*, 13(9), e0203958.

Torres, R. R., Gerhart, N., & Negahban, A. (2018). Combating fake news: An investigation of information verification behaviors on social networking sites. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (pp. 3976-3985). Honolulu: HICSS.

Tseng, E. (2015). *Providing relevant notifications based on common interests in a social networking system*, U.S. Patent No. 9,083,767. Washington, DC: U.S. Patent and Trademark Office.

Volkova, S., & Jang, J. Y. (2018). Misleading or falsification: Inferring deceptive strategies and types in online news and social media. In *Companion Proceedings of the Web Conference 2018* (pp. 575-583). Geneva: International World Wide Web Conferences Steering Committee.

Vydiswaran, V. G. V., Zhai, C. X., Roth, D., & Pirolli, P. (2012). BiasTrust: Teaching biased users about controversial content. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management* (pp. 1905-1909). New York: ACM.

Vydiswaran, V. G. V., Zhai, C. X., Roth, D., & Pirolli, P. (2015). Overcoming bias to learn about controversial topics. *Journal of the Association for Information Science and Technology*, 66(8), 1655-1672.

Walter, S., Brüggemann, M., & Engesser, S. (2018). Echo chambers of denial: Explaining user comments on climate change. *Environmental Communication*, 12(2), 204-217.

Zimmer, F. (2019). Fake news. Unbelehrbar in der Echokammer? In W. Bredemeier (Ed.), *Zukunft der Informationswissenschaft. Hat die Informationswissenschaft eine Zukunft?* (pp. 393-399). Berlin: Simon Verlag für Bibliothekswissen.

Zimmer, F., Akyürek, H., Gelfart, D., Mariami, H., Scheibe, K., Stodden, R., ... Stock, W. G. (2018). An evaluation of the social news aggregator Reddit. In *5th European Conference on Social Media* (pp. 364-373). Sonning Common: Academic Conferences and Publishing International.

Zimmer, F., & Reich, A. (2018). What is truth? Fake news and their uncovering by the audience. In *5th European Conference on Social Media* (pp. 374-381). Sonning Common: Academic Conferences and Publishing International.

Zimmer, F., Scheibe, K., Stock, M., & Stock, W. G. (2019). Echo chambers and filter bubbles of fake news in social media. Man-made or produced by algorithms? In *8th Annual Arts, Humanities, Social Sciences & Education Conference* (pp. 1-22). Honolulu: Hawaii University.

Zimmer, F., Scheibe, K., & Stock, W. G. (2018). A model for information behavior research on social live streaming services (SLSSs). In Meiselwitz G. (Ed.), *Social Computing and Social Media. Technologies and Analytics: 10th International Conference, Part II* (pp. 429-448). Cham: Springer.

Zuckerberg, M., Sanghvi, R., Bosworth, A., Cox, C., Sittig, A., Hughes, C., ... Corson, D. (2006). *Dynamically providing a news feed about a user of a social network*, U.S. Patent No. 7,669,123 B2. Washington, DC: U.S. Patent and Trademark Office.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice
eISSN : 2287-4577 pISSN : 2287-9099

Research Paper
J Inf Sci Theory Pract 7(2): 54-64, 2019
https://doi.org/10.1633/JISTaP.2019.7.2.5

A Theory of Public Knowledge

Shawne D. Miksa*
Department of Information Science, College of Information, University of North Texas, Denton, TX, USA
E-mail: Shawne.Miksa@unt.edu

Chinami McLain
Department of Information Science, College of Information, University of North Texas, Denton, TX, USA
E-mail: ChinamiMcLain@my.unt.edu

ABSTRACT
A theory of public knowledge is offered for the purposes of defining more clearly its role in information systems and classification schemas. Public knowledge is knowledge intended to be available for use in a public system. It is knowledge accessible to the public or knowledge in the public arena as opposed to the other seemingly multitudinous ways to describe knowledge. Furthermore, there are many different public arenas or small worlds. Public knowledge, irrespective of these different arenas, has four important overlying characteristics: It is consensual, it does not imply complete truth or certainty, it is autonomous, and it has a constant renewal of old knowledge with new knowledge. Each of these attributes has been culled from a study of the works of Patrick Wilson, Karl Popper, and John Ziman.

Keywords: public knowledge, information systems, classification, theory, small worlds

Open Access

Received date: May 23, 2019
Accepted date: June 08, 2019

***Corresponding Author:** Shawne D. Miksa
Associate Professor
Department of Library and Information Sciences, College of Information, University of North Texas, 3940 N Elm St, Denton, TX 76207, USA
E-mail: Shawne.Miksa@unt.edu

1. INTRODUCTION

The idea of public knowledge remains a general idea, not something easily seen through to something real and tangible. It is a convenient idea that rings true—that is, people do organize themselves into groups, or disciplines…which can be described as small worlds of people with shared social values and norms who produce knowledge. Disciplines, in particular, are important components of public knowledge. It is a convenient idea in the sense that speaking of knowledge as a *universe of knowledge* is a convenient idea. It is difficult to prove the existence of a universe of knowledge, but the idea that knowledge is somehow interrelated as a single whole is eminently sensible. Another similarly convenient idea is that of saying that groups of people know things. An individual person can know things. However, in what sense does a group of people know anything? There is no group brain, in the same sense that individual brains are the bases of individual knowing. There are only individual brains, in fact. Nevertheless, it is commonly accepted that people together in a group know things; that there exists knowledge of the public arena. Car drivers commonly know that red traffic lights mean *stop* and green lights mean *go*, bakers know that most bread requires flour of some type, and everyone who has ever been in an ocean knows the water is salty. Alternatively, from a point of view of a discipline, archaeologists have specific knowledge about ancient cultures gained through the study of found artifacts from those cultures, generating theories about how their societies were organized or traded with other societies, how they spread out over territories, and so on. How does such a phenomenon take place?

Social or collective knowledge can be spoken of as group knowledge based on an analogy from individual knowing. In short, groups of people know in the same sense that an individual knows something, but that statement is analogical—and in the end only a convenient way of speaking. What is the mechanism by which a group of people—a small world—knows anything? Is it simply accepted as a reality because is it sensible to believe that many different groups of people do in fact find common knowledge within their group?

However, the question becomes more complex—a complex knot of individual threads. If public knowledge is assumed to be comprised of knowledge coming from hundreds and hundreds, perhaps millions, of these small worlds then how is a particular 'sphere of knowledge' identified? What would be its characteristics? Outward evidence? How is it measured? What are its operational definitions such that it could confidently be said that it is legitimate? It is one thing to posit their existence, and quite another to say positively they have been 'seen.'

On a more practical note, however, it seems logical to say that an understanding of public knowledge is vital to library and information science, especially as it pertains to access and retrieval of the information resources produced by individuals and groups of people that constitute a large part of the universe of knowledge. Swanson (1986) explored the role and importance of information retrieval 'facilitating' discoveries of scientific knowledge, focusing specifically on knowledge that goes 'undiscovered.' Similarly, the construction and modification of the classification schema (e.g., Dewey Decimal Classification [DDC], Universal Decimal Classification, etc.) should also recognize the distinctive characteristics of public knowledge. A model or theory of public knowledge might be a good first step in defining its role in information systems and classification schema.

This is a purely theoretical paper intended to examine what is meant by 'public knowledge' or 'knowledge accessible to the public,' or 'knowledge in the public arena' as opposed to other seemingly multitudinous ways to describe 'knowledge.' What is this special type of knowledge and why is there is a need to distinguish it from other types of knowledge? Moreover, having defined public knowledge, what is its role in information retrieval systems or knowledge classification schemas? If an information system (a broad term here) is created to organize and collocate public knowledge, then defining that knowledge and recognizing that the knowledge may be unique to a specific group of users is necessary.

2. DEFINING PUBLIC KNOWLEDGE

Public knowledge is knowledge intended to be available for potential use in a public system which is any system that is available to a group of people. It could also be called social knowledge or collective intelligence but with the caveat that there is not one collective intelligence or one type of social knowledge. Public knowledge is not *subjective or personal* knowledge, which, by definition, falls outside the realm of public knowledge. Therefore, this special kind of knowledge will be offered here as a set of concepts and proposition statements that describe the characteristics of this type of knowledge. The concepts are numerous but a refined set of the more prominent ones can be assembled such as knower, knowledge, consensus, truth, autonomy, ownership, and those concepts that hint at the cyclical nature of public knowledge such as interaction, new knowledge, and old knowledge. Propositional statements follow:

- Public knowledge is consensual knowledge.
- Public knowledge does not imply truth.

- Public knowledge is autonomous in the sense that it may or may not require a knower. It does not belong to any one person and it may be undiscovered.
- New public knowledge is produced to replace the old knowledge that no longer explains phenomena.

One last clarification is needed before discussing more deeply these four propositional statements. The phrase 'knowledge of the public arena' is not intended to mean the public at large, or all publics. Rather, it is the *small worlds* that individuals may inhabit together. Chatman (1999) defined small worlds as

> …a society in which mutual opinions and concerns are reflected by its members, a world in which language and customs bind its participants to a worldview. Resources (both intellectual and material) are known and easily accessible. It is a world in which there is a collective awareness about who is important and who is not; which ideas are relevant and which are trivial; whom to trust and who to avoid. In its truest form, a small world is a community of like-minded individuals who share co-ownership of social reality (p. 213).

Chatman's definition is used here instead of, for example, Jaeger and Burnett's (2010) broader definition of an 'information world.' One person may inhabit many small worlds, worlds that have definite boundaries from others. Differences in academic disciplines is a good example—the arena of a humanities scholar is different in many ways from the mathematical arena but at the same time they both function in the larger small world of a university or academia overall. In today's U.S. political climate, the stark differences in beliefs about the direction of the country's future can be keenly felt, as is evidenced by debates on globalization, climate change, immigration, gender fluidity, and socio-economic differences. The U.S. is one country, but with an infinite number of small worlds. In a way, this is an exploration of public knowledge when seen in the context of a small world. These characteristics of public knowledge have been culled from an examination of a number of published works. In particular, Patrick Wilson's *Public knowledge, private ignorance*, John Ziman's *Reliable knowledge*, and Karl Popper's *Objective knowledge* were examined. Wilson was a professor of library and information science who wrote on bibliographic organization, Ziman a theoretical physicist who wrote extensively on the matter of the social nature of science and scientific knowledge, and Popper was a critical thinker and philosopher of science, well-known for his work on objective knowledge. There are many others that could be examined, perhaps most notably would be Swanson's (1986) work on undiscovered public knowledge and Polanyi's (1962) work on personal knowledge. Swanson's (1986) idea of 'undiscovered public knowledge' was influenced by Karl Popper's Third World—the world of objective knowledge, or recorded knowledge. He wrote that "the world of published knowledge certainly contains more than any of person can know and indeed contains more than the aggregate of what all persons know" (p. 107).

Wilson (1977) addressed what is needed for an effective library and information policy and in doing so concluded that an effective policy "should be based on an understanding of the way in which what is known is represented in, and recoverable from, the documents that constitute the library's chief, though not sole, stock" (p. vii). Ziman (1978) addressed the reliability of scientific knowledge by showing that it is "a product of a collective human enterprise to which scientists make individual contributions which are purified and extended by mutual criticism and intellectual cooperation" (p. 3). Lastly, Popper (1979) raised interesting questions on the nature of objective knowledge which has had a lasting influence on several thinkers within the library and information sciences, among other disciplines.

Popper's Third World, or the world of objective knowledge, is the realm of the "objective contents of thought, especially of scientific and poetic thoughts and of works of art" (Popper, 1979, p. 106). Scientific thought includes theoretical systems, critical arguments, problems, and problem situations. Poetic and artistic works, while having originated in the subjective mind, become objective when presented publicly. They may be reinterpreted at various times, but the essence of the work remains unchanged. This Third World includes all the contents of all the libraries (p. 107). Most importantly, Popper insisted that this world exists independently of the human mind; in other words, it is totally independent of anyone's claim to know and it is autonomous in the sense that it can be acted upon by us, but it cannot be mastered by us (p. 112). The idea of public knowledge existing independently of the mind is challenging; public knowledge once recorded may separate itself from its creators, but it needs users to exist and to grow. It does not hover or float above us literally, although today people often put their information in 'the cloud' (i.e., server farms) figuratively.

Ziman's (1978) notion of scientific knowledge is somewhat different from Popper's Third World. He described scientific knowledge as being *intersubjective*; in other words, it can only be validated and translated into action by the intervention of human minds. In contrast to Popper's insistence that the Third World exists independently of human minds, Ziman wrote "that it does not contain any independent source of objective knowledge" (p. 8). Scientists unavoidably interject their world

view into their work and it is impossible to be unbiased in their scientific observations because of this world view. He also wrote that scientific knowledge is consensual because "individual contributions are purified and extended by mutual criticism and intellectual cooperation" (p. 3). This process of criticism and cooperation is the "scientific paradigm" and constitutes how the objective knowledge of the scientific realm comes into being (p. 7).

Ziman and Popper both explore the arenas of the physical, the subjective, and the objective. Popper calls them simply World One, World Two, and World Three whereas Ziman refers to them as simply the material domain, the mental domain, and the noetic domain, or the world of the objective: a world that is "collectively created and maintained as a social institution" (Ziman, 1978, p. 106). Ziman in turn took this idea of the noetic from Polanyi (1962), who wrote of "superior knowledge" as that which is "coherently believed to be right and excellent by men within their culture" (p. 375).

Wilson (1977) does not specifically use the term 'objective' nor does he focus particularly on scientific knowledge. Rather, he is addressing these concepts through the study of what he calls 'public knowledge.' In this sense he sees it as something that is decided on by the public itself. Public knowledge is the "best constructed view of the world at any given time, judged by our own best procedures for criticism and evaluation" (p. 5). It is not only created and accepted as a special type of knowledge, but at the same time people evaluate it and judge it to the best presently available story (for lack of a better term). This type of knowledge includes much that is not known. It can include, but is not limited to, scientific knowledge. In that sense, Wilson does imply that public knowledge is objective and scientific.

3. THE UNIVERSE OF KNOWLEDGE AND PUBLIC KNOWLEDGE

Earlier it was postulated that speaking of knowledge as a *universe of knowledge* is a convenient idea and that while it is difficult to 'prove' the existence of a universe of knowledge, the idea that knowledge is somehow interrelated as a single whole is eminently sensible. The intent of this paper is to provide a theory of public knowledge which can be seen as a part of the larger universe of knowledge. The idea of a universe of knowledge is discussed further on in this paper but suffice it to say that it may help to clarify the purpose of this paper by making clear the relationship between the universe of knowledge and public knowledge. One way is to describe it is as a whole-to-part relationship. Public knowledge (whole) is objective knowledge that can consist of various individual knowledge (subjective knowledge) parts. And, the universe of knowledge (the larger whole) consists of various public knowledge (parts).

4. ATTRIBUTES OF PUBLIC KNOWLEDGE

4.1. Consensual Knowledge

Knowledge in any small world is knowledge that is agreed upon; it comes from the general agreement of the majority of the population of the small world. What is known at the moment is called upon to fulfill information needs until what is known changes or is replaced because it does not satisfy the information need. This process takes place in both scientific and non-scientific small worlds. Ziman and Wilson both explicitly state that knowledge formation is a consensual process. In science, the goal "is to achieve the maximum degree of consensuality" (Ziman, 1978, p. 6). Wilson sees it as the public agreeing on what is, or what is not, a relevant or acceptable part of the consensus. In other words, people have control over the decision of what is or what is not to be public knowledge. Ziman and Wilson share the view that not only does society create and accept public knowledge, but it also decides upon the criteria for evaluation and judgment. In this sense, knowledge formation is an ongoing process of negotiation. Similarly, Swanson (1986) sees the necessity of public criticism by scientists; published arguments create the published products that make up public knowledge. More recently, Ma (2015) has argued that although group consensus alone does not make knowledge, "the construction of public knowledge is inarguably social and discursive" (p. 535).

Popper (1979) implies human consensus but he posits that once knowledge is proposed it is 'out there' and becomes part of something that humans no longer have much control over. He writes that "all work in science is work directed towards the growth of objective knowledge" and that this growth takes place by the "formulation of problems, the emergence of new problem situations, competing theories, mutual criticism by way of argument: all these are the indispensable means of scientific growth" (pp. 121-122). Yet, he writes that contributions by man to this objective knowledge are "vanishingly small" and that no one can master any one part of it (p. 161). The formation of knowledge is a process with an end goal—growth—and while negotiation, or consensus, is part of the process is more a super-consensual process.

4.2. Truth

The second characteristic of public knowledge is that it does not imply truth. In other words, while much of what is accepted as public knowledge will be believed and used by the public as

foundations for creating new knowledge or discrediting old knowledge, everything is subject to change as beliefs themselves change or because not all the knowledge is known; as Swanson (1986) would say, some public knowledge is *undiscovered*.

The aim of science is to find "*satisfactory explanations* of whatever strikes us as being in need of explanation" and these explanations are "more or less well known to be true, or assumed to be so known" (Popper, 1979, p. 191). However, Popper dismisses *ultimate explanations* which he defines "an explanation which is neither capable of any further explanation, nor in need of it" (p. 194). Instead, he contends that every explanation may be "further explained by a theory or conjecture of a higher degree of universality" (p. 195). Put another way, for every explanation given more explanations are generated, and therefore the essential nature of something, the *ultimate essence*, can never be fully described (p. 196).

Ziman (1978), speaking from within the same scientific realm as Popper (1979), approaches this idea of truth from a different angle. He does not believe there are absolute truths because the scientific paradigm "does not contain any independent source of objective knowledge" (p. 8). There is no truly objective knowledge because a scientist, despite a rigorous training process when studying to become an expert in any given field, cannot contradict his own world view with "statements that are obviously at variance with what he has learnt and come to love" (p. 8). Scientists cannot truly divorce their status as human beings from the unbiased observer that is the ideal scientist. Ziman also questions whether the scientific paradigm is any more believable as a unique world picture than any other world view such as that held by a non-scientific group. While both Popper and Ziman agree that public knowledge in a scientific community does not implicitly hold truth, they come to their conclusions from different avenues. Popper obviously believes scientists are truly unbiased. Ziman contends they can never be so because they are human.

Wilson (1977) does not limit his discussion of public knowledge to the scientific realm and he uses another approach. He stresses that his explanation of public knowledge is not to be taken as an explanation of all knowledge; public knowledge is simply the public stock of knowledge at any given time and can include things that are not known to anyone. But, when looking at the standard concept of knowledge he writes, it is "at the very least true belief, and without this belief there would be no knowledge (p. 6). *True belief* is not the same as *truth*, however. It simply signifies the complete acceptance of a piece of knowledge until, at some future time, that belief in the truth is shifted to another piece of knowledge. Said another way, knowledge, according to Wilson, implies belief and truth in something that is known and since public knowledge may include things that are not known, it cannot be truly *believed* and so cannot be true. "Certainty", he writes, "has no dominant role in the theory of public knowledge public knowledge no more implies certainty then it implies truth" (p. 6). This could be seen also as a consequence of there never being a complete world view. It may contain "vague and indefinite views" or be altogether blank in some areas (p. 5).

Public knowledge is defined as knowledge held by a specific group of people who, by consensus, agree what the knowledge will be, and also understand that this special knowledge will not inspire complete certainty or complete belief in its truth due to the fact that is it constantly changing or may contain knowledge that is not known. This leads in turn to the third and fourth features of public knowledge.

4.3. Known Versus Not Known

Wilson's (1977) use of 'known' versus 'what is not known' reveals the third characteristic of public knowledge: that it is potentially autonomous. Knowledge may or may not require a knower. Wilson asserts that public knowledge includes much that is not known to anyone; that some knowledge may be unknown for years until it is rediscovered. Some knowledge may not be adequately captured in published documents, and some may only be passed on verbally, but even then it may not be completely understood (p. 9). The technological expertise needed to discover the knowledge may even be lacking—whether intelligent life exists off-world or the cure for diseases such as HIV or Alzheimer's. Either way, making knowledge public does not guarantee that someone will acquire it or understand it, in other words that they acquire knowledge (p. 9). It may 'float' on the public arena of the small world but when it is accessed or abstracted by someone it becomes subjective at that particular instance.

Ziman and Popper both contend that public knowledge, or objective knowledge, does not need a knower to be knowledge. Popper (1979) writes

> Knowledge in this objective sense is totally independent of anybody's claim to know; it is also independent of anybody's belief, or disposition to assent; or to assert, or to act. Knowledge in the objective sense is *knowledge without a knower*: it is *knowledge without a knowing subject* (p. 109).

His declaration here relies strongly on his belief that objective knowledge, what he calls the Third World, is autonomous; it exists independently of the human mind. Popper asserts that a book is still a book even if it is never read. Everything contained

within the book is objective knowledge, as if no more than extant messages. Whether anyone reads it or not, whether it is true or false, whether anyone really understands it, does not change its status as objective knowledge. Furthermore, it is the *potentiality* of someone understanding, or even misunderstanding, the message that makes everything contained in the book objective knowledge (p. 116). However, Popper does concede that "a book should—in principle, or virtually—be capable of being grasped (or deciphered, or understood, or 'known') by somebody. But I do not admit more" (p. 116).

Ziman, too, holds this view. In fact, he directly quotes Popper when discussing objective knowledge (Ziman, 1978, p. 107). His understanding relies on the fact that scientific knowledge does not come from any one individual but is the result of the cooperation of many individuals. The final product, he says, belongs "to humanity" (p. 108). He writes that this is true of non-scientific knowledge as well – be it artistic knowledge, social knowledge, etc. Beyond quoting Popper, Ziman does not have as strong an argument that knowledge does not require a knower to be considered knowledge. He offers the explanation that our belief in scientific knowledge comes mainly from our accepting it because "eventually it gives its own semblance to our picture of the world" (p. 108). People come to accept the objectivity of knowledge not necessarily by knowing it personally, but by simply accepting it because it was accepted as knowledge by a consensus of the public.

Whether or not knowledge needs a knower to be knowledge is debatable when considering the views of these three writers. On the one hand, knowledge may not exist unless someone understands it, or it may exist even if it is never discovered, or even if it is misunderstood.

4.4. New Knowledge Replaces Old Knowledge

Public knowledge has so far been described as a consensual, not necessarily true, autonomous entity that may or may not require a knower. Ziman, Popper, and Wilson all agree that new knowledge is constantly being produced to replace the old knowledge that no longer adequately explains phenomena. Ziman (1978) writes that "much of the research literature of science is intended *rhetorically* – to persuade other scientists of the validity of a new hypothesis or to shatter received opinions" (p. 7). This is necessary in a scientific community in order to ensure the maximum amount of consensus and it is accomplished by "mutual criticism and intellectual cooperation" (p. 3).

Popper (1979), holding to the view of an autonomous, objective world of knowledge, sees the growth of new knowledge resulting from the "feed-back effect" that objective knowledge has on those who would use it (p. 161). Referring back to his non-belief in absolute truths, every explanation of a phenomena creates more questions and more explanations. This is a necessary condition for the growth of knowledge and as a result there will always be "an infinity of problems" that will remain undiscovered (p. 161). Similarly, Wilson asserts that public knowledge has to be *constructed*. He writes that "we have again and again to survey the state of knowledge, or the state of the different areas" and that this is the "job of construction" (Wilson, 1977, p. 10). There is a constant cycle of criticism and evaluation to ensure that pubic knowledge is the best view of the world at any given time (p. 5).

In short, public knowledge develops from the consensus of a group of people with similar interests or goals; it may or may not be certain or true, and it may not be known to all; and it seems to be in a constant state of change or adjustment.

5. PUBLIC KNOWLEDGE AND INFORMATION SYSTEMS

In library and information science public knowledge is addressed at the level of controlling the information resources. This is commonly called 'bibliographic' control but it involves more than just books, so it is appropriate to use the phrase 'information resources' or 'information-bearing entities', as coined by F. Miksa. People strive to provide organization of, and access to, the information that is produced through information retrieval systems with the idea of providing a means of communication between the users of the systems and the information resources in the system, or even the system itself. And, just as there are many types of knowledge there are also many types of information systems employed in the task of providing access to and retrieving ordered knowledge. The next logical step is to identify the defining characteristics of the system that would make public knowledge accessible and retrievable. And, looking further, how should public knowledge be organized within this information system?

How can one define the essential components of a public knowledge information system? This will not be a technical review of the electronic innards or the programming complexities of such a system. Rather, it is about the conceptual elements with which the system should be composed. Once again, this will be culled from the work of three individuals: Michael Buckland's *Information and information systems*, Daniel Bell's *The coming of the post-industrial society*, and Wilson's *Two kinds of power: An essay on bibliographic control*. Buckland (1991) focuses exclusively on the "nature

of information systems" and not on the technical aspects. Wilson (1977) discusses the specifications of 'bibliographical instruments' from which we hope to extrapolate to an information system that uses all types of information resources. Bell (1973) offers some notions of technology that may serve to fine-tune the overall definition being sought.

Starting with Bell (1973), it is possible to take a step back and examine the idea of an information system from a broader view. Bell discusses the notion of intellectual technologies and writes that one of the major problems with the post-industrial society will be the management of large-scale systems, with large numbers of interacting variables, which have to be coordinated to achieve specific goals" (p. 29). This is easily found daily in the online interactions with search engines and information systems, as well as social media, online shopping, the 24-hour news cycle, and wearable devices with applications that send data constantly to 'the cloud.' Bell defines an intellectual technology as the substitution of algorithms (problem solving rules) for intuitive judgments" (p. 29). Algorithms are used because it is the nature of complex systems to be counter-intuitive; in other words, there are too many variables interacting for our minds to hold and process successfully. The cause and effect relationships in a complex system may be too "deeply hidden or remote in time, or may lie in the very structure (i.e., pattern) of the system itself" (p. 32). Naturally, computers are used to run these algorithms. However, putting aside an exploration of algorithms, it is more important to explore what sort of conceptual framework the algorithms should be employed in so as to provide order over a constantly changing arena of knowledge.

Buckland (1991) views information systems as depending on information processing—"deriving new forms and representations of existing information" (p. 28). They are open systems, not isolated from social and technological contexts. The system is likely to be large and complex in terms of its elements and the relationships involved, such as social, economic, political, and cognitive activities. It will also respond to changes and will adapt itself to environments in order to survive.

These complex, open, adaptable systems are also contingent upon responses—responses that "constitute the means of change and adaption by internal alteration, by changing relationships, or by influencing the external environment" (Buckland, 1991, p. 28). Buckland names five responses: inquiry, perception, becoming informed, demand, and provision. Information must be perceived to be communicated, observed, or retrieved. Inquiry arising from perception is a motivation to know something and this motivation will shape the use of the information system. Demand for information arises from the perception of information, as does inquiry. The provision of information is motivated by the "goals, preferences, and perceptions by those who have resources that can be allocated for that purpose" (p. 29). As individuals become informed by the information received, or perceived, they can then inquire about and receive even more information. These responses can be viewed as a continuous cycle and perhaps as a finer way of stating what Bell (1973) called the cause-and-effect relationships in complex systems.

Buckland (1991) distinguishes between information systems that communicate or observe and those that are retrieval-based which involve "the additional complexity of selecting, collecting, retrieving, and searching information (p. 30). The role and mission of the retrieval-based system is to facilitate access to information and to support the mission of whoever finds the information—those individuals who are members of many small worlds. This can be related back to the public arenas of knowledge discussed earlier. The role and mission of each of these arenas will certainly vary and as such demand a unique information retrieval system.

By information Buckland (1991) writes of *information-as-thing*—the physical artifacts, the data, and the documents that people see as information resources or information-bearing entities. An information system deals directly with this type of information. Now, having discussed public knowledge and its characteristics—some of which are not tangible—it is not possible to create a retrieval system based only on tangibles. And, by that what is meant is something beyond born-digital information resources—something recorded and created as a 'document,' for all intents and purposes. The knowledge that is 'out there' is not tangible, but it is known in some capacity. In the next section on classification schemas Buckland's notion of *information-as-knowledge* and its relationship to these schemas as they are used within information systems will be discussed, but for now the focus is on systems dealing with tangibles.

There are other factors to consider when making these systems. Wilson (1968) describes five elements in the specification of what he calls 'bibliographic instruments' which can be related to information systems. Wilson defines bibliographical instruments as that which "consists entirely or primarily of descriptions of works, texts, and copies" (p. 57). The five specifications are the rules for the construction of the instrument:

- domains of the instrument—the set of items of which the system will consist, including those items that may be considered for inclusion
- principles according to which the items represented have been drawn—what claims can be made about the domain

- determinations of what is to count as a unit for listing and description—knowing by what rules an item has been determined as 'listable'
- what information can be expected to be found about an item when it is listed as a unit
- the complex system of arrangement or organization—where an item of a given sort will be found and what it means to find an item in a given place (pp. 59-62)

Wilson writes that these specifications alone do not ensure the success of an instrument. It must be known how well they have been followed by the designers of the instrument. By specifying the domain of items, the designers guarantee the items will meet the requirements for inclusion as well as the principles used to decide the requirements. If these principles are not discoverable then no claims can be made about the domain. If the user does not understand the "often quite complicated rules by which it is determined what is a 'listable unit,' one is likely to make the grossest of errors in using the instrument" (p. 61). Also, to include representations of the items, as surely must happen because it is not always possible to include the actual artifact, the user must be made aware of what information he or she will find as well as what information is not found. Lastly, knowing how the organization or arrangement of the items is implemented in the system allows the user to immediately and directly identify "items that fit some description without the necessity of scanning all the descriptions of the items listed" (p. 62).

Information systems have been described as being complex intellectual technologies that are not wholly separated from the small worlds they are serving. These systems not only house information but process it as well. This processing is contingent upon the responses of the system user. Lastly, the user must be able to understand the specifications by which items have been collected, represented, and organized within the system.

6. PUBLIC KNOWLEDGE AND CLASSIFICATION SCHEMA

This raises the question of how to build a classification system based on the public knowledge of a small world. Is it a viable option to make ultra-specialized classification schemas? Should we? We have attempted to describe public knowledge not as a universal public knowledge but public knowledge that is a natural outcome of a small world in which individuals "share co-ownership of the reality of that world" (Chatman, 1999, p. 213). Library classification schema used in most academic libraries attempt to organize what is considered 'all' of knowledge (i.e., arts and humanities, physical sciences, social sciences, etc.) and as such may not be as specific as is needed to meet information needs. There have been studies throughout the library and information science (LIS) field, especially in its formative years of the mid-twentieth century, looking at the different needs of scholars in different fields of research, and some have concluded that these differing information needs call for information systems that take these differences into account.

Library classifications have largely centered around the concept of the *universe of knowledge* which in found in the early writings of H. E. Bliss, E. C. Richardson, W. C. Berwick Sayers, and S. R. Ranganathan, and those who studied under them. The undertaking of library classification invariably demands an understanding of all that can potentially be housed in a library setting—that being, at the highest level, all of humankind's knowledge. Many in LIS refer to this as the *universe of knowledge: Universe* because it includes everything and *knowledge* because people see that product of human minds as one of the most valuable possessions besides, perhaps, life itself. As such, it rightfully belongs in a library. The *universe of knowledge* is essentially a limitless area of all our knowings, and because people are in constant need of this knowledge they devise ways of finding and retrieving the 'containers' of knowledge, the information resources and all they contain, or in some systems the knowledge itself. This notion is a direct product of the beginnings of the modern library, starting in the nineteenth century. A good public library collection should strive to have the 'best'—the best books by the best authors, representing all the main divisions of humankind's recorded knowledge.

There was also the belief that organization of knowledge should be based on the organization of the sciences. Bliss, Richardson, Berwick Sayers, and Ranganathan all held that the organization of the sciences was the cornerstone of what is considered *knowledge*. This can be seen most strongly in the work of Richardson and Bliss. In Richardson's *Classification: Theoretical and practical* he devotes his first lecture to the order of the 'sciences,' rather than saying the order of 'knowledge.' His whole argument is based on the premise that "the order of things is the order of the sciences" (Richardson, 1930, p. 10). Things, from his viewpoint, are anything that have separate existence: anything that is, in other words (pp. 1-2).

Bliss (1929) writes that the unification of knowledge is "one of the highest purposes of science and philosophy" (p. 166). He distinguishes science from other types of knowledge because he believes our classifications should be scientific in their orientation as science is reflective of the order of nature. Science, he writes, "is verified and organized knowledge, experiential,

rational, methodic, proceeding from generalizations, theories, and conceptual systems" (p. 190). The universe is made up of, and relies on, the interweaving of relationships between things, concepts, and classes and relations themselves and science, with philosophy, is dedicated to revealing these relationships (p. 165). Thus, the ordering of things that comes from the sciences reveals a true order, or a natural order. Richardson and Bliss greatly rely on what they term Nature. Nature, to Bliss, is the "system of real things and relations external to human minds and underlying the works of humanity" (Bliss, 1929, p. 173). Nature, to Richardson (1930), is that which is outside of man and that which is already classified (p. 2). In their natural order there is a natural order in the universe and as science inherently seeks to reveal this natural order, so it must be the purveyor, the champion so to speak, of the arrangement of the universe of knowledge.

Berwick Sayers (1938) seems to assume, without too much introspection, what these two men are saying as he cites them often in his own work. He takes his stance by expounding specifically on book classification and how it should be accomplished. He asks if there is an order in nature and, if so, should classification follow it? He gives no definite answer but rather turns to the 'logical machinery' that should drive a classification scheme (p. 10). Further on he makes the simply statement that classification of books is more the classification of the statements contained within their pages (p. 16). He might say that most books deal with something of a scientific nature; therefore a practical classification, recognizing the nature of books, is more scientifically-based. In his introduction he says as much when he writes:

> My classification theory is quite simple. The order which philosophers, scientists, or valid systematic thinkers have discovered in things is the basis of book classification (Berwick Sayers, 1938, p. xix).

Ranganathan takes a different approach to the situation, owing in large part to his work having come in the later years following Richardson, Bliss, and Berwick Sayers, as well as due to his differences in his education compared to the education of these three men.[1] He never completely or directly states that classification should follow the scientific order. Interestingly, he was first and foremost a mathematician before he ever set foot in the realm of LIS. His idea of faceted classification is heavily influenced by his prior mathematical training, and it should be noted that he developed his colon classification prior to actually developing his theory of library classification (Miksa, 1998, p. 1). When Ranganathan talks of applying his classification scheme his examples tend toward products of the scientific realm. For example, at the time he wrote, he believed classification had the most to offer in this realm. Certainly, in the present time, there is even more information being produced and more and more scientific specializations being practiced. At one point he likened the influx of articles written and published to a swarm of locusts (Ranganathan, 1967, p. 206). In his view, the realm of subjects was multi-dimensional and could be seen in the scientific world as it was developing before his very eyes in the mid-twentieth century.

How does this idea of the organization of knowledge being based on the organization of science impact the theory of public knowledge as expressed so far? Scientific disciplines, as small *worlds*, share some of the four characteristics discussed in the first half of this paper. In fact, "we see Science... as public knowledge at its most manifest" (Ziman, 1968, p. 53). However, not every arena of public knowledge is scientifically-based and not all information resources housed in libraries and classified with current classification schema are derived in purely scientific ways. For examples, books on witchcraft are certainly not considered scientific but in the DDC they are split between the 100s (Philosophy and Psychology), the 200s (Religion), the 700s (Arts), and the 800s (Literature and Rhetoric). Melvil Dewey built his classification on the principle that knowledge could be divided into ten main divisions, while at the same time intending it to be an "open and expanding system" (Miksa, 1987, p. 7). The DDC system has since developed into a very elaborate knowledge organization system (currently in its 23rd edition), but it still relies on the same ten main divisions. The Library of Congress Classification was built upon the idea of literary warrant and the categories created varied according to the "goal of arranging the subjects of each area of knowledge in a unique and tailored manner" (Miksa, 1984, p. 29). This certainly holds true to the idea of new knowledge replacing old knowledge, but it still tries to encompass a *universe of knowledge* and so has become quite a cumbersome system.

This is not to quibble that classification schema should not be ordered as science. However, instead of molding public knowledge to conform to what may be a limited view would it not be better to let the nature of public knowledge guide its own organization?[2] Beghtol (2003) described a cyclical relationship

[1] See, for example, F. Miksa's The *DDC, the universe of knowledge, and the postmodern library* (1998), and his 1997 paper on the influence of mathematics on Ranganathan's work.

[2] While a discussion of folksonomy and crowdsourcing of information does have some bearing on this topic, our perception is that folksonomies have become rather passé in the LIS literature and crowdsourcing of information projects such as Wikipedia have had to rely more and more on strict editing and verification of information; all of which is food for thought for another paper.

between naïve classification schemas (new classification schema invented for purposes of knowledge discovery) and classification schema traditionally used in information retrieval systems. The primary difference between these two types is that naïve schemas are generated when discovering unrecorded knowledge (i.e., scholarly activity and research) which is then disseminated through publications, which in turn are classified as documents so as to be made discoverable by users searching information retrieval systems. Beghtol also points out the relationship between 'literary warrant' and 'consensus' in scholarly activity—as publications of new research are disseminated among scholars, literary warrant becomes extensive and thus consensus develops among scholars (p. 70). In that sense, new knowledge is constantly being generated and replaces old knowledge.

Of course, it is possible to just dismiss the whole idea as being too general by pointing out that different groups of people or different disciplines still have different modes of operations and different ways of seeing relationships and connections, and so on. That, in fact, attempting to chart public knowledge is an enormous task, perhaps impossible, for more than one reason and that it is not adequate for information retrieval or classification schema due mainly to its generality and constantly shifting nature. But, would that not then be a good reason to use one classification schema so as to *relate* and connect all the different spheres together?

7. CONCLUSION

This discussion started by suggesting that an understanding of public knowledge is vital to library and information science, especially as it pertains to access and retrieval of the information resources produced by individuals and groups of people (who may belong to many small worlds) and that constitute a large part of the *universe of knowledge*. Similarly, the construction and modification of the classification schema that would seek to provide order over this type of knowledge should recognize the distinctive characteristics of public knowledge. This would increase the chances of not only meeting the information needs of users but also of creating more cohesive specialized knowledge systems for future use.

We have offered a theory of public knowledge such that:

- Public knowledge is consensual knowledge.
- Public knowledge does not imply truth.
- Public knowledge is autonomous in the sense that it may or may not require a knower. It does not belong to any one person and it may be undiscovered.
- New public knowledge is produced to replace the old knowledge that no longer explains phenomena.

Public knowledge is created by groups of individuals in a process that is never ending—much like Otlet's (1990) notion of a universal book of knowledge that "will never be completed but will grow unceasingly" (p. 84). There is constant interaction between individuals and public knowledge, just as in user-centric information systems there is constant interaction between the system and individuals. As noted earlier, Richardson believed that the order of things in nature should be mirrored in the order of the sciences; classifications are still, in many ways, our (human) interpretation of that order in nature. We are trying to show our understanding of nature/world with a collection of public knowledge created by various groups of individuals. Public knowledge (whole) is objective knowledge that consists of various individual knowledge (subjective knowledge) parts. And the universe of knowledge (larger whole) consists of various public knowledge (parts). Within each whole-part relationship, there is constant interaction between the whole and the parts, just as there is also constant interaction between the universe of knowledge and individuals.

REFERENCES

Beghtol, C. (2003). Classification for information retrieval and classification for knowledge discovery. *Knowledge Organization*, 30(2), 64-73.

Bell, D. (1973). *The coming of the post-industrial society: A venture in social forecasting*. New York: Basic Books.

Berwick Sayers, W. C. (1938). *An introduction to library classification* (5th ed.). London: Grafton.

Bliss, H. E. (1929). *The organization of knowledge and the systems of the sciences*. New York: Holt.

Buckland, M. (1991). *Information and information systems*. Westport: Praeger.

Chatman, E.A. (1999). A theory of life in the round. *Journal of the American Society for Information Science*, 50(3), 207-217.

Jaeger, P. T., & Burnett, G. (2010). *Information worlds: Social context, technology, and information behavior in the age of the Internet*. New York: Routledge.

Ma, L. (2015). A sign on a tree: A case for "public knowledge. *Library Trends*, 63(3), 528-539.

Miksa, F. (1984). The *development of classification at the Library of Congress* (Occasional Paper). Urbana: University of Illinois, Graduate School of Library and Information Science.

Miksa, F. (1987). The influence of mathematics on the classificatory thought of S.R. Ranganathan. In *Knowledge Organization for Information Retrieval: Proceedings of the Sixth International Study Conference on Classification Research*, University College, London, 16-18 June, 1987 (FID 716, pp. 167-179). The Hague: International Federation for Information & Documentation.

Miksa, F. (1998). *The DDC, the universe of knowledge, and the post-modern library*. Albany: Forest Press.

Otlet, P. (1990). *International organization and dissemination of knowledge: Selected essays of Paul Otlet* (W. B. Rayward, Trans.). Amsterdam: Elsevier.

Polanyi, M. (1962). *Personal knowledge: Towards a post-critical philosophy*. Chicago: University of Chicago Press.

Popper, K. (1979). *Objective knowledge: An evolutionary approach* (Rev. ed.). Oxford: Clarendon Press.

Ranganathan, S. R. (1967). *Prolegomena to library classification* (3rd ed.). New York: H. W. Wilson.

Richardson, E. C. (1930). *Classification: Theoretical and practical* (3rd ed.). New York: H. W. Wilson.

Swanson, D. R. (1986). Undiscovered public knowledge. *The Library Quarterly*, 56(2), 103-118.

Wilson, P. (1968). *Two kinds of power: An essay on bibliographical control*. Berkeley: University of California Press.

Wilson, P. (1977). *Public knowledge, private ignorance: Toward a library and information policy*. Westport: Greenwood Press.

Ziman, J. (1968). *Public knowledge: The social dimension of science*. Cambridge: Cambridge University Press.

Ziman, J. (1978). *Reliable knowledge: An exploration of the grounds for belief in science*. Cambridge: Cambridge University Press.

Call for Paper

Journal of Information Science Theory and Practice (JISTaP)

We would like to invite you to submit or recommend papers to **Journal of Information Science Theory and Practice** (JISTaP, eISSN: 2287-4577, pISSN: 2287-9099), a fast track peer-reviewed and no-fee open access academic journal published by Korea Institute of Science and Technology Information (KISTI), which is a government-funded research institute providing STI services to support high-tech R&D for researchers in Korea. JISTaP marks a transition from Journal of Information Management to an English-language international journal in the area of library and information science.

JISTaP aims at publishing original studies, review papers and brief communications on information science theory and practice. The journal provides an international forum for practical as well as theoretical research in the interdisciplinary areas of information science, such as information processing and management, knowledge organization, scholarly communication and bibliometrics.

We welcome materials that reflect a wide range of perspectives and approaches on diverse areas of information science theory, application and practice. Topics covered by the journal include: information processing and management; information policy; library management; knowledge organization; metadata and classification; information seeking; information retrieval; information systems; scientific and technical information service; human-computer interaction; social media design; analytics; scholarly communication and bibliometrics. Above all, we encourage submissions of catalytic nature that explore the question of how theory can be applied to solve real world problems in the broad discipline of information science.

Co-Editors in Chief : Gary Marchionini & Dong-Geun Oh

Please click the "Online Submission" link in the JISTaP website (http://www.jistap.org), which will take you to a login/ account creation page. Please consult the "Author's Guide" page to prepare your manuscript according to the JISTaP manuscript guidelines.

Any question? Ji-Young Kim (managing editor) : jistap@kisti.re.kr

Information for Authors

The Journal of Information Science Theory and Practice (JISTaP), which is published quarterly by the Korea Institute of Science and Technology (KISTI), welcomes materials that reflect a wide range of perspectives and approaches on diverse areas of information science theory, application and practice. JISTaP is an open access journal run under the Open Access Policy. See the section on Open Access for detailed information on the Open Access Policy.

A. Originality and Copyright

All submissions must be original, unpublished, and not under consideration for publication elsewhere. Once an article is accepted for publication, all papers are accessible to all users at no cost. If used for other researches, its source should be indicated in an appropriate manner and the content can only be used for uncommercial purpose under Creative Commons license.

B. Peer Review

All submitted manuscripts undergo a single-blind peer review process in which the identities of the reviewers are withheld from the authors.

C. Manuscript Submission

Authors should submit their manuscripts online via Article Contribution Management System (ACOMS). Online submission facilitates processing and reviewing of submitted articles, thereby substantially shortening the paper lifecycle from submission to publication. After checking the manuscript's compliance to the Manuscript Guidelines, please follow the "Online Submission" hyperlink in the top navigation menu to begin the online manuscript submission process.

D. Open Access

With the KISTI's Open Access Policy, authors can choose open access and retain their copyright or opt for the normal publication process with a copyright transfer. If authors choose open access, their manuscripts become freely available to public under Creative Commons license. Open access articles are automatically archived in the KISTI's open access repository (KPubS, www.kpubs.org). If authors do not choose open access, access to their articles will be restricted to journal users.

E. Manuscript Guidelines

Manuscripts that do not adhere to the guidelines outlined below will be returned for correction. Please read the guidelines carefully and make sure the manuscript follows the guidelines as specified. We strongly recommend that authors download and use the manuscript template in preparing their submissions.

1. Page Layout :
All articles should be submitted in single column text on standard Letter Size paper (21.59 × 27.94 cm) with normal margins.

2. Length :
Manuscripts should normally be between 4,500 and 9,000 words (10 to 20 pages).

3. File Type :
Articles should be submitted in Microsoft Word format. To facilitate the manuscript preparation process and speed up the publication process, please use the manuscript template.

4. Text Style :
- Use a standard font (e.g., Times New Roman) no smaller than size 10.
- Use single line spacing for paragraphs.
- Use footnotes to provide additional information peripheral to the text. Footnotes to tables should be marked by superscript lowercase letters or asterisks.

5. Title Page :
The title page should start with a concise but descriptive title and the full names of authors along with their affiliations and contact information (i.e., postal and email addresses). An abstract of 150 to 250 words should appear below the title and authors, followed by keywords (4 to 6).

Author1
Affilliation, Postal Address. E-mail

Author2
Affilliation, Postal Address. E-mail

ABSTRACT
A brief summary (150-250 words) of the paper goes here.

Keywords : 4 to 6 Keywords, separated by commas.

6. Numbered Type :

1. INTRODUCTION
All articles should be submitted in single column text on standard letter size paper (21.59 × 27.94 cm) with normal margins[1 . Text should be in 11-point standard font (e.g., Times New Roman) with single line spacing.
[1 Normal margin dimensions are 3 cm from the top and 2.54 cm from the bottom and sides.

2. SECTIONS

The top-level section heading should be in 14-point bold all uppercase letters.

2.1. Subsection Heading 1

The first-level subsection heading should be in 12-point bold with the first letter of each word capitalized.

2.1.1. Subsection Heading 2

The second-level subsection heading should be in 11-point italic with the first letter of each word capitalized.

7. Figures and Tables :

All figures and tables should be placed at the end of the manuscript after the reference list. To note the placement of figures and tables in text, "Insert Table (or Figure) # here" should be inserted in appropriate places. Please use high resolution graphics whenever possible and make sure figures and tables can be easily resized and moved.

Figure

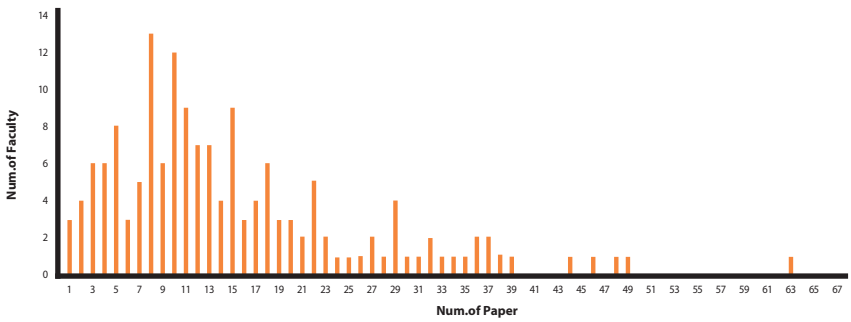


Fig. 1. Distribution of authors over publication count.

Table

Table 1. The title of table goes here

Study	Time period study	Data
Smith Wesson (1996)	1970 - 1995	684 papers in 4 SSCI journals
Reeves [a (2002)	1997 - 2001	597 papers in 3 SSCI journals
Jones Wilson [b (2011)	2000 - 2009	2,166 papers in 4 SSCI journals

[a Table footnote a goes here
[b Table footnote b goes here

8. Acknowledgements :

Acknowledgements should appear in a separate section before the reference list.

9. Citations :

Citations in text should follow the author-date method (authors' surname followed by publication year).

- Several studies found... (Barakat et al., 1995; Garfield, 1955; Meho & Yang, 2007).
- In a recent study (Smith & Jones, 2011)...
- Smith and Jones (2011) investigated...

10. Reference List :

Reference list, formatted in accordance with the American Psychological Association (APA) style, should be alpha-betized by the first authors last name.

Journal article

- Author, A., Author, B. & Author, C. (Year). Article title. *Journal Title*, volume(issue), start page-end page.
- Smith, K., Jones, L. J., & Brown, M. (2012). Effect of Asian citation databases on the impact factor. *Journal of Information Science Practice and Theory*, 1(2), 21-34.

Book

- Author, A., & Author, B. (Year). *Book title*. Publisher Location: Publisher Name.
- Smith, K., Jones, L. J., & Brown, M. (2012). *Citation patterns of Asian scholars*. London: Sage.

Book chapter

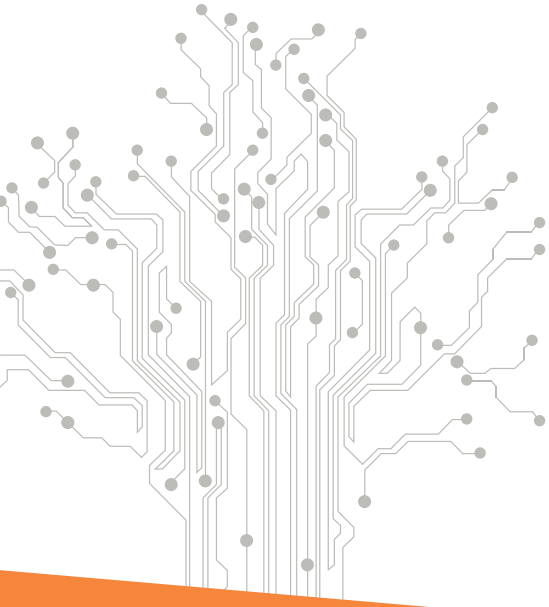
- Author, A., & Author, B. (Year). Chapter title. In A. Editor, B. Editor, & C. Editor (Eds.), *Book title* (pp. xx-xx). Publisher Location: Publisher Name.
- Smith, K. & Brown, M. (2012). Author impact factor by weighted citation counts. In G. Martin (Ed.), *Bibliometric approach to quality assessment* (pp. 101-121). New York: Springer.

Conference paper

- Author, A., & Author, B. (Year). Article title. In A. Editor & B. Editor (Eds.), *Conference title* (pp. xx-xx). Publisher Location: Publisher Name.
- Smith, K. & Brown, M. (2012). Digital curation of scientific data. In G. Martin & L. J. Jones (Eds.), *Proceedings of the 12th International Conference on Digital Curation* (pp. 41-53). New York: Springer.

Online document

- Author, A., & Author, B. (Year). Article title. Retrieved *month day*, *year* from URL.
- Smith, K. & Brown, M. (2010). The future of digital library in Asia. *Digital Libraries*, 7,111-119. Retrieved *May 5, 2010*, from http://www.diglib.org/publist.htm.



JISTaP

Journal of Information Science
Theory and Practice

http://www.jistap.org

KISTI
www.kisti.re.kr

66, Hoegi-ro, Dongdaemun-gu, Seoul, Republic of Korea (ZIP code: 02456)
Tel. +82-2-3299-6102 Fax. +82-2-3299-6067 http://www.jistap.org