



Review

# Supporting Newsrooms with Journalistic Knowledge Graph Platforms: Current State and Future Directions <sup>†</sup>

Marc Gallofré Ocaña \* and Andreas L. Opdahl

Department of Information Science and Media Studies, University of Bergen, 5020 Bergen, Norway; andreas.opdahl@uib.no

\* Correspondence: marc.gallofre@uib.no

† This paper is an extended version of our paper published in Proceedings of the CIKM 2020 Workshops.

**Abstract:** Increasing competition and loss of revenues force newsrooms to explore new digital solutions. The new solutions employ artificial intelligence and big data techniques such as machine learning and knowledge graphs to manage and support the knowledge work needed in all stages of news production. The result is an emerging type of intelligent information system we have called the Journalistic Knowledge Platform (JKP). In this paper, we analyse for the first time knowledge graph-based JKPs in research and practice. We focus on their current state, challenges, opportunities and future directions. Our analysis is based on 14 platforms reported in research carried out in collaboration with news organisations and industry partners and our experiences with developing knowledge graph-based JKPs along with an industry partner. We found that: (a) the most central contribution of JKPs so far is to automate metadata annotation and monitoring tasks; (b) they also increasingly contribute to improving background information and content analysis, speeding-up newsroom workflows and providing newsworthy insights; (c) future JKPs need better mechanisms to extract information from textual and multimedia news items; (d) JKPs can provide a digitalisation path towards reduced production costs and improved information quality while adapting the current workflows of newsrooms to new forms of journalism and readers' demands.

**Keywords:** journalistic knowledge platform; artificial intelligence; knowledge graph; intelligent information system; newsrooms; journalism



**Citation:** Gallofré Ocaña, M.; Opdahl, A.L. Supporting Newsrooms with Journalistic Knowledge Graph Platforms: Current State and Future Directions. *Technologies* **2022**, *10*, 68. <https://doi.org/10.3390/technologies10030068>

Academic Editor: Manoj Gupta

Received: 10 April 2022

Accepted: 26 May 2022

Published: 31 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

News agencies and news organisations are under pressure from the loss of advertisement and revenues [1,2], and facing an audience that is less likely willing to pay for digital content [3,4]. Despite an increase in digital consumption, information is no longer consumed from a limited number of TV stations and news outlets. Instead, readers have access to and can contrast fresh and first-hand information from free-available sources on the internet and social media at any time. As a consequence of their freedom of choice, readers demand high-quality journalism [5] and trusted sources [4,6,7].

In response, news agencies and news organisations are constantly adapting their business models to digital media innovations in order to improve information quality, competitiveness and growth [8]. Innovation and digitalisation of newsrooms are needed to increase the quality and lower the cost of news production, changing how journalists and readers interact with news content and background information [9]. Newsrooms are therefore embracing big data and artificial intelligence (AI) techniques such as knowledge graphs and machine learning (ML) for journalistic purposes [10,11] such as identifying and contextualising newsworthy events in investigative journalism; facilitating data visualisation in digital journalism; analysing information in data journalism; automating news writing in robot journalism; providing real-time fact-checking tools for political journalism. The result is an emerging type of intelligent information system that we call the *Journalistic*

*Knowledge Platform (JKP)* which is currently gaining interest in research and practice. In this paper, we define JKPs as platforms that apply AI and big data to journalism in order to manage and support the knowledge work needed in all stages of news production.

JKPs can be described from a functional, an organisational and a technical perspective. From a functional point of view JKPs automate the process of annotating metadata and support daily workflows like news production [12,13], archiving [14,15], management [16,17] and distribution [18–21]. JKPs harvest and analyse news and social media information over the net in real time [22], leverage encyclopaedic sources [23], and provide journalists with both meaningful background knowledge [24] and newsworthy information [25]. From an organisational viewpoint: JKPs are deployed in newsrooms to manage the knowledge needed to support journalists with creativity and discovery tasks. These are tailored to the particular digital strategies and editorial lines to improve news broadcast. JKPs also follow media standards to facilitate communication with customers and providers, and are subject to legal regulations such as data privacy. From a technical perspective JKPs implement state-of-the-art AI technologies such as machine learning, natural language processing (NLP) and knowledge representation and reasoning. News-relevant information is represented in knowledge bases which are exploited with data analysis, reasoning and information retrieval techniques to help journalists and readers dive more deeply into information, events and storylines. Today, knowledge graphs [26] are a topical technique for knowledge representation that continues to grow in importance, therefore, we centre our analysis on JKPs building on knowledge graphs.

According to the authors of Hogan et al. [26], knowledge graphs capture and abstract knowledge using graph-based data models. They are particularly relevant for scenarios that integrate and extract value from diverse and dynamic data. Wherein entities of interest are represented as nodes and the relations between them as edges of the graph. Ontologies and rules are used to define the semantics and terms of the graph and reason about it, but also to ease data integration from, for example, Linked Open Data (LOD) [27] and existing large-scale knowledge graphs like Wikidata and DBpedia. Compared to relational and NoSQL models, knowledge graphs facilitate semantic integration, flexible data and schema evolution and graph query languages with mechanisms to explore complex relations through arbitrary-length paths.

In this article, we explore the current state and suggest future research directions for knowledge graph-based JKPs. We ask: “What challenges and opportunities for newsrooms have motivated the knowledge graph-based JKPs?” (RQ1), “How does the research on knowledge graph-based JKPs address these challenges and opportunities?” (RQ2) and “What are the most important open areas for research on knowledge graph-based JKPs?” (RQ3). To answer these three questions we have performed a detailed analysis of 14 JKPs reported in the literature that apply AI and big data to journalism in order to manage and support the knowledge work needed in all stages of news production. A broader literature on related technologies exists. Our analysis does not ignore other solutions applying artificial intelligence to journalism, but our focus is on providing a comprehensive analysis of the main concepts of those JKPs that build on knowledge graphs rather than specific techniques, optimisations, tools and systems. The JKPs were selected in context of a broader systematic literature review on how knowledge graphs can support news in a wide sense [28]. Compared to this study, Opdahl et al. [28] was not restricted to JKPs and did not analyse the challenges and opportunities nor the current and future directions of JKPs. We conducted a qualitative meta-analysis (see Appendix A for a detailed description of the meta-analysis method), and we examined the existing JKPs in light of our experiences with developing JKPs along with an industry partner for the international newsroom market [29].

The present article extends Gallofré Ocaña and Opdahl [30], which analyses challenges and opportunities for developing JKPs along six axes: stakeholders, information, functionalities, techniques, components and concerns. This article extends the analysis by considering more JKPs. It investigates how well the challenges and opportunities are cov-

ered in the research literature and suggests future research directions. The rest of the paper is organised as follows: we summarise the identified JKPs in Section 2; analyse the current challenges and opportunities for newsrooms that motivated JKPs in Section 3; present the state of research on JKPs in terms of their stakeholders, information, functionalities, techniques, components and concerns in Section 4; discuss the future directions for research on JKPs in Section 5.

## 2. Analysed Platforms

We identified 14 platforms that fit under our definition of JKPs, which we list in Table 1. The identified JKPs cover a total of 28 papers carried out by distinct research groups located in 11 different countries and in collaboration with a variety of news agencies, news organisations and industry partners.

**Table 1.** Selected platforms. N: news media partner and T: technology partner. The identified countries represent the news media partners' countries.

Platform	Industry partners	Countries	References
PlanetOnto	-	UK	[18,31]
Neptuno	Diari SEGRE <sup>N</sup> and iSOCO <sup>T</sup>	Spain	[14]
AnnoTerra	NASA's Earth Observatory <sup>N</sup>	USA	[24]
SemNews *	-	USA	[19]
Hermes *	-	The Netherlands	[20,32,33]
BBC CMS	BBC <sup>N</sup>	UK	[16,34]
NEWS	Agencia EFE <sup>N</sup> , Agencia ANSA <sup>N</sup> and Ontology Ltd. <sup>T</sup>	Spain and Italy	[12,35]
EventRegistry *	-	Slovenia	[21]
NewsReader *	LexisNexis <sup>T</sup> , The Sensible Code Company (before ScraperWiki) <sup>T</sup> and Synerscope <sup>T</sup>	Netherlands, Spain and Italy	[15,36,37]
Reuters Tracer	Reuters <sup>N</sup>	USA	[22,38,39]
SUMMA	LETA <sup>N</sup> , BBC Monitoring <sup>N</sup> , Deutsche Welle <sup>N</sup> and Priberam Labs <sup>T</sup>	Latvia, UK, Germany	[17,40–42]
INJECT	Adresseavisen <sup>N</sup> , AFP <sup>N</sup> , The Globe and Mail <sup>N</sup> , Stibo <sup>T</sup>	Norway, France, Canada	[13]
ASRAEL	AFP <sup>N</sup>	France	[23]
News Hunter ‡	Wolftech <sup>T</sup>	Norway	[29,43,44]

‡ News Hunter is the JKP in which the authors are involved. \* Related systems that can be used as JKPs—either directly or with some adaptations—but have not been published in the context of newsrooms.

The JKPs from 2000 to early 2010 implemented the Semantic Web idea [45] in newsrooms. These JKPs used semantic web technologies [46] to automate the metadata annotation process [16], combine different knowledge bases [24], and formalise media standards [14]. They used ontologies in NLP pipelines together with Linked Open Data (LOD) [27] resources from external knowledge bases (i.e., Wikipedia, DBpedia) to automatically annotate news archives and feeds with metadata about topics, keywords, categories and other relevant information (e.g., persons, places, organisations, sentiments and relations) [14,19]. The annotated information was stored in knowledge bases, facilitating the interlinking of news across different archives, online catalogues and external LOD repositories [16,24]. For instance, Neptuno was the first project to publish a journalistic

ontology and adapt the IPTC topics [47] as RDF [14,48], and Troncy [49] converted the IPTC NewsCodes [50] into SKOS [51] thesaurus and defined an OWL [52] ontology for the IPTC News Architecture [53]. The resulting systems provided services for supporting news creation [14], personalising news retrieval [18,20], facilitating semantic search [14,18,19,24,33], visualising ontologies [14], managing content [16], aggregating information [24,34] and recommending news [16].

The JKPs from early 2010s until today focused on identifying and analysing events and advancing AI/ML for supporting journalism. In addition, some of them focused on scaling over large volumes of live streams of multimedia news [36], social media [39] and TV/radio broadcasts [17]. Similar to the previous JKPs, news items were annotated using either media standards [35] and LOD resources [37] or both [23] and stored in knowledge bases to facilitate cross-lingual information retrieval services through semantic technologies and ontologies [21,23]. These JKPs continuously monitored and curated the annotated items using AI/ML and LOD to provide relevant insights for journalists and identify current, past and future events. For example, the annotated news items were used to identify networks of actors [15], suggest news angles [13,54], automate news creation [22] and facilitate fact-checking [17], and the events were analysed using different AI/ML techniques for grouping events and news items [21,23], reasoning over events, and reconstructing the evolution of the events along time [15].

### 3. Challenges and Opportunities Facing Newsrooms

In current newsroom workflows, metadata annotation like tagging and categorisation is often performed manually by journalists. This is a time-consuming process that is error-prone, imprecise and restricts future usability [12]. The added metadata is reduced to a few general categories that are limited to authorship, dates, content language and news management information. This metadata is used to address newsworthiness and filter events according to news customers' and audiences' interests. However, due to the lack of fine-grained annotations, newsrooms have difficulties implementing high-quality information retrieval and filtering services [14,16,20]. Hence, they return irrelevant, incomplete and even biased results to customers [21].

Journalists spend a lot of their time monitoring and filtering large volumes of news feeds like TV broadcasts, radio shows, social media and published news to keep them up-to-date, time that otherwise would have been invested in producing news [42]. Today's worldwide daily news volumes scale over 100,000 articles making it unfeasible for journalists to manually handle tasks like fact-checking and searching for related articles. Germann et al. [41] (p. 1) claim that "each of [BBC] ca. 300 monitoring journalists usually keeps track up to 4 live sources in parallel (typically TV channels received via satellite), plus a number of other sources of information such as social media feeds". This is an undesired situation for a business sector where time is a critical factor, delays can lower the value of information and imply economic losses [35].

This massive volume of textual and multimedia data is often organised in different catalogues or databases and managed by external services [24,35]. Because these catalogues are not integrated nor share a common schema and lack fine-grained annotations, they limit the possibilities for newsrooms to extract valuable insights and knowledge. Structuring the information and integrating the data from a variety of sources bring newsrooms with better ways to exploit data and facilitate the adoption of AI. For example, it can ease the implementation of information retrieval services and recommender systems and the automation of news creation processes and the detection of fake news and newsworthy events.

To help with these processes, newsrooms currently use a mix of proprietary systems, external services, tools and in-house taxonomies or categorisation schemas that are challenging to integrate and operate together [35,55]. It is a complex ecosystem of applications that hinders the expansion and evolution of digitally integrated newsrooms. It makes it difficult for managers to get an overview of what is happening in news rooms [41]. It limits the interaction with customers [35]. Additionally, it can lead to vendor binding or dependence

situations due to the difficulties of maintaining multiple and diverse proprietary solutions. All together and with the urge of reducing cost, increasing high-quality journalism and adapting current newsrooms to digital advances, journalists and newsrooms are becoming interested in the services that JKPs can offer [9].

#### 4. State of Research on JKPs

We describe the state of research on JKPs by investigating the stakeholders, information, functionalities, techniques, components and concerns dealt with in the identified JKPs. These six analysis axes are based on a qualitative analysis reported in an earlier paper [30].

##### 4.1. Stakeholders

JKPs provide services to and interact with a large variety of stakeholders. Figure 1 shows the identified stakeholders and their three top-level categories: general user, organisation and technical agent.

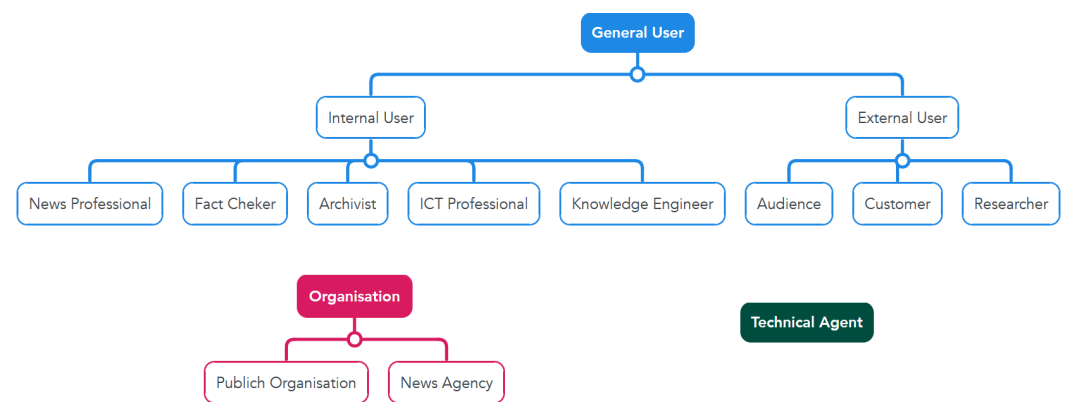


Figure 1. Stakeholder categories.

The general users can be divided between the internal users that belong to newsrooms and the external ones. The internal users are news professionals like journalists who use JKPs for creating histories [35,39]; fact-checkers who conduct an essential task in combating with fake news and misinformation [17]; archivists who maintain up-to-date the schemas and news archives [14]; ICT professionals and knowledge engineers who develop and maintain JKPs [12]. Whereas, the external users are the audience [21]; the customers to whom new agencies offer services and researchers who investigate JKPs or use JKP to analyse data, as in the SUMMA project where “[political scientists want] to perform data analyses based on large amounts of news reports” [42] (p. 2).

JKPs support organisations in different ways: The most direct is in news agencies and news organisations where JKPs are deployed and adapted to particular digital strategies and purposes, but also to other news organisations that consume services from external JKPs. Moreover, JKPs provide services to both private and public organisations like governmental agencies that interact with or consume services from newsrooms, for example, the SUMMA project “provides media monitoring and analysis services to [...] the British government” [42] (p. 1). JKPs also interact indirectly with the organisations responsible for controlling news media standards, vocabulary and ontologies (e.g., the IPTC organisation). This impacts how JKPs are designed because the work of many news agencies depends on those standards, and JKPs often need to build on and comply with them. However, the media standards may not cover or fit the use cases of newsrooms, as in the NEWS project where “most of the NewsCodes defined by IPTC do not have alternative versions in different languages, only in English” [35] (p. 9). Hence, JKPs need to adapt or expand the media standards according to their needs.

Last but not least, the technical agent represents the JKPs and any system or technical infrastructure in newsrooms that support or interact with JKPs. A sub-type of the technical



agent is the external system that communicates with newsroom services, like the customers' information systems [35].

#### 4.2. Information

JKPs cover the whole news production pipeline from gathering information and news creation to knowledge exploitation and distribution. Table 2 lists the identified categories of information.

**Table 2.** The most common types of information managed by JKPs.

Information	Explanation
News content	The reported story or event.
Textual data	Textual information.
Multimedia data	Images, videos and audio information.
Data format	The format in which the data is stored or structured.
Metadata	Data about or that describe the news content.
Linked Open Data (LOD)	Structured and open available data on the Internet (e.g., data from Wikidata and DBpedia) [27]
Events	Newsworthy happenings.
Information needs	Different information types and categories of interest.

JKPs deal with textual and multimedia news content produced by news agencies, news organisations and external sources that are managed and distributed to customers and audience [12,14,15]. As textual data we consider the raw text from any source like news articles, social media feeds, web pages, blogs, PDF files, biographies, reports, historical data and geopolitical data. Whereas, as multimedia we consider live broadcasts, photographs, audio files and video files. Moreover, news agencies produce and distribute content in different formats like plain text, Information Interchange Model (IIM), News Industry Text Format (NITF), NewsML and RDF [16,35].

News content is annotated and enriched with metadata using LOD, semantic vocabularies and ontologies, for example, the ASRAEL project “leverage[s] the Wikidata knowledge base to produce semantic annotations of news articles” [23] (p. 1). Metadata can describe different types of basic information like the authorship, language, creation time, ownership, media type, priority, status, version, keywords and categories; as well as inferred information like provenance, tone and sentiment, and the relevant persons, stories, locations, organisations and events [14,34,37].

Journalists and customers of newsrooms are highly interested in current events and their related information [12]. In addition, JKPs are designed to support additional information needs: General users want to have access to details about the stories (i.e., who, what, why, where and when), identify networks of actors and implications, search the events based on their type or place, obtain facts, and retrieve evidences [15,16,24]. News professionals need access to news archives and knowledge bases for documentation purposes, finding connections from past events, following histories and identifying emerging topics [14,35,36,42]. Additionally, customers have different information needs depending on their business or interests, for example, “the press cabinet of a company is usually interested in news items talking about the company or its rivals, whereas a sports TV channel is interested mostly in news items describing sports events” [35] (p. 1).

#### 4.3. Functionalities

JKPs provide different functionalities to their users. Table 3 lists the identified main functionalities.

**Table 3.** Most common type of functionalities and services provided in JKPs.

Functionality	Explanation
News creation	The process to create a news story.
Verification	The process of checking the facts and claims.
Source selection	The ability to select the information sources of interest.
Monitoring	The ability to continuously distil information from source.
Knowledge discovery	Functionalities for exploring relevant information.
Trends	The current newsworthy developments.
Alert	A notification.
Summarisation	Extracting and representing the key information from a larger text or group of text.
Clustering	Grouping similar stories or events.
Business support	Functionalities to support management workflows.
Content management	Functionalities oriented to store, organise and distribute information.
Personalisation	Providing information according to the user's interests.

News professionals use JKPs for news creation. This creative process involves different tasks such as discovering, collecting, organising, contextualising and publishing [56,57]. JKPs guide news professionals in writing up their stories [29], support them with contextual background knowledge [12,13,29], provide the means for comparing current events with other events [23] and facilitate access to previous work for creating similar content for a different audience, region or language [42]. JKPs also support news professionals with verification [58] tasks like fact-checking [19,59], provenance [15], rights and authorship management [35]. These are typically time-consuming tasks for journalists and fact-checkers that JKPs automate [17].

Source selection and monitoring functionalities are common across the studied JKPs that harvest and store content from internal and external sources and monitor them in real-time [19,21,36,42]. These functionalities allow journalists to automatically follow and distil news and social media of interest and relieve them from these time-consuming tasks.

Knowledge discovery [60] is one of the most attractive functionalities of JKPs. It allows users to obtain news insights, analysis and relevant information. For instance, in NewsReader it “increases the user understanding of the domain, facilitates the reconstruction of news story lines, and enables users to perform exploratory investigation of news hidden facts” [15] (p. 1). Other interesting functionalities among the studied JKPs are the trends identification used to discover emerging topics, long-term developments and changes in events over time [21,37]; alerts to keep users up-to-date with the last incoming items [19,31,41]; summarisation [61] of news histories and events to provide additional insights [21]; clustering of story lines and events [23,42].

JKPs can be used as business support systems to manage and monitor internal newsrooms production, news coverage and broadcast decisions [31,42]. This helps managers and editors in allocating resources, avoiding duplicate work and detecting news that can be relevant to different audiences. JKPs are also used for content management that allows newsrooms to store, organise and distribute the daily produced content and meta-data [14,16,35].

Most of these functionalities should be personalised and tailored to the stakeholders' needs. Hence, JKPs allow the personalisation of their functionalities according to users' preferences and profiles [12,18,33].

#### 4.4. Techniques

JKPs implement and combine different IT techniques to fulfil their functionalities. Table 4 lists the IT techniques that we have identified.

**Table 4.** The most common IT techniques used in JKPs.

Technique	Explanation
Semantic technologies	Set of technologies designed to work with LOD and semantic data [46].
Fact extraction	The techniques used to identify factual claims.
Conceptual model	A representations of the world or a part of.
Reasoning	The techniques used to infer knowledge.
Network analysis	The techniques used to analyse networks of things.
Event analysis	The techniques used to analyse events.
Natural Language Processing (NLP)	A set of techniques intended to work and process language.
AI training	The process of creating and tuning an AI model to perform on a given dataset or scenario.

Semantic technologies [46] and similar semantic representation techniques are widely utilised in all the studied JKPs. They use semantic technologies for automating annotation, disambiguating, enriching and leveraging news items with information from external knowledge bases [12,14,19,37]. The semantic representations provide neutral language, explicit relations and facilitate structural matching and lingual independence. They are used for clustering news items and events [23] and detecting trends and story lines [15]. These semantic representations together with fact extraction techniques are used to obtain factual claims from news items and link them to their sources and facts in external knowledge bases (e.g., Wikidata, Wikipedia) [15,19,42].

Conceptual models provide vocabularies, schemas and ontologies. These are often implemented using semantic technologies and represent news stories, events and related information. In addition, conceptual models can define users' interests and preferences [18,20,35], and provide shared resources and formats to facilitate content management and semantic interoperability [14,16,24,37].

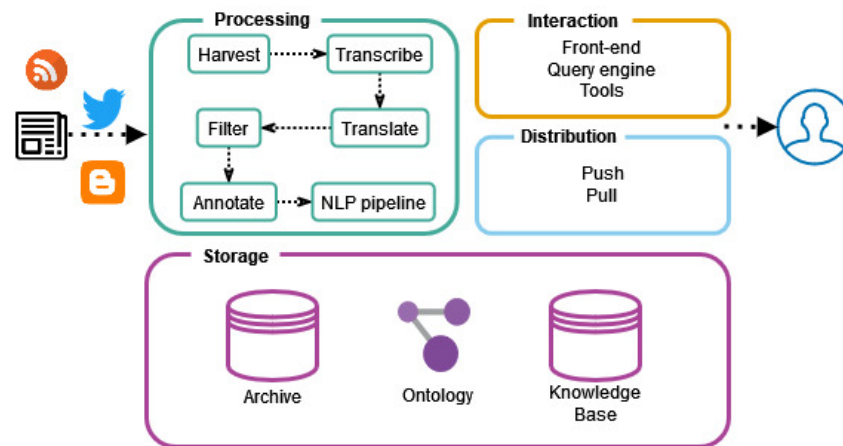
Conceptual models and semantic technologies are also used for reasoning, network analysis and event analysis. Reasoning techniques abstract and infer new knowledge from news items, events and temporal aspects [37]. Network analysis is used to find networks of actors, organisations and their implications [15]. Event analysis is applied to detect, identify, cluster and annotate the events described in the news [21,23,35].

The aforementioned techniques are supported by NLP tasks such as named entity recognition, relation extraction and temporal expression normalisation [19–21,37,40]. These NLP tasks, among others, are used in many of the components and functionalities of JKPs. In order to obtain optimal results from the NLP tasks, near-continuous training on extensive news corpora [23] is needed to always keep the machine learning models up-to-date.

#### 4.5. Components

JKPs rely on different components to fulfil their functionalities and support users. We split these components into four groups: processing, storage, interaction and distribution (see Figure 2). The processing components deal with harvesting data from different sources and processing them. The storage components store and manage data. The interaction components allow users to interact with the information from the system and the distribution components distribute information to users.





**Figure 2.** JKP components.

The processing components cover tasks from data gathering to transforming input sources into knowledge representations. The textual and multimedia sources are continuously harvested. However, not all contents receive the same interest from news professionals, like in SUMMA where “entertainment programming such as movies and sitcoms, commercial breaks, and repetitions of content (e.g., on 24/7 news channels) [...] [are] of limited interest to monitoring operations” [42] (p. 1). Thus, the harvested content is also translated [42] and filtered according with the different stakeholders’ interests and needs. In the studied JKPs, spoken content is transcribed [42] and images are textually described [12] to be able further process them.

The harvested content is automatically annotated with metadata (e.g., authorship, categories and topics) to support functionalities like business support, content management and personalisation [14,31,33,35]. The annotated content is often processed by a NLP pipeline using state-of-the-art NLP and natural language understanding modules to perform linguistic tasks such as co-reference resolution, named entity recognition, relation extraction and sentiment analysis [15,19,62]. Both the results of the NLP pipeline and the annotated content are represented semantically following a predefined schema or ontology. These representations link the annotations to a knowledge base (e.g., an RDF-based knowledge graph) [20,37] and enrich the news items with facts from external knowledge bases (e.g., the LOD cloud, DBpedia and Wikidata) [15,23].

The storage infrastructure of a JKP can be composed of an archive, an ontology and a knowledge base. The archive can store millions of historical news articles, biographies, reports [14,37] and other relevant textual and multimedia items. The knowledge base is where the annotated semantic representations of news items are stored and enriched with external information [14,15,24]. The ontology is used to represent the structure of the news items, leveraged information, metadata and vocabulary [14,24,31,35]. Most recent JKPs also include dedicated storage for real-time news-related feeds [42].

Stakeholders interact with the previous components and have access to the functionalities of JKPs mainly by using three types of interaction components: front-ends that implement specific functionalities, for example, news editors with automatic annotation for creating news articles, statistical and visual analysis features for generating reports [19,21] and enhanced insights for discovering new stories [18]; tools that provide useful resources for creating news like currency converters and dictionaries [35]; query engines that can be accessed through APIs and user interfaces. These allow journalists and customers to query, explore, analyse and visualise the archives and knowledge bases [16,19,20,31,42].

News agencies and news organisations use the push and pull components for delivering and distributing content to their users. Push components offer interfaces where information consumers can select and subscribe to feeds of news [12,16,19,31,41]. Whereas the pull components are used to access and browse the repositories of JKPs [14,16,21,31,35].

#### 4.6. Concerns

Stakeholders, information, functionalities, techniques and components are influenced or affected by additional concerns of various types. Table 5 lists the identified concerns.

**Table 5.** Concerns related to JKPs.

Aspect	Explanation
Customers heterogeneity	The diversity of newsroom customers.
Standards	Standards like IPTC topics or RDF.
Ownership	Copyrights, authorship and licensing information.
Multilingual content	Content produced in various languages.
Timeliness	The temporal aspect of news, when they are published and when the stories happen.
Human factors	Human-related aspects that affect newsroom and JKPs.
Quality	The information and data quality.
Big data	Aspects related to the large volume of data, variety of data and velocity in which data is produced.
Performance	The ability to provide results with the expected quality and on time.
Legacy	Old systems or repositories.
Software architecture	The structure and components of a software system [63].
Maintenance	The ability to reuse, fix and update existing systems.

The customers of JKPs are heterogeneous. They cover diverse sectors and industries, from other newsrooms to companies and institutions, and use different systems to interact with JKPs [35,42]. To improve the interoperability between news agencies and stakeholders, JKPs utilise standards like the IPTC news codes, media topics, semantic vocabularies and RDF [14,35], and keep track of information related to ownership, such as authorship, copyrights, privacy and sources [12,64]. JKPs can also use the ownership information to control the information provenance and reliability [15] by, for example, tracking back the information to its original source and identifying trustworthy providers.

Customers and audiences prefer different languages [21,23,35,37,42]. Hence, JKPs deal with and produce multilingual news items (e.g., Norwegian, Italian, Spanish, English) that are translated, transcribed and delivered in the preferred languages. In addition, these news items have an intrinsic timeliness aspect that defines their value either as a fresh event or as part of a past or present storyline or historic development that can be reconstructed [12,15,20,42].

JKPs attempt to address different human factors in newsrooms. JKPs automate error-prone and time-consuming processes that were performed manually like news tagging, source monitoring, information filtering, verification, fact-checking and finding related articles and relevant information [14,17,19,21,35]. Hence, JKPs free journalists from these tedious tasks and improve their results. As a result, JKPs facilitate high-quality information to meet the standards of their stakeholders [12].

On the technical side, JKPs deal with big data requirements like volume, velocity, variety. ASRAEL estimates that “the number of collected articles ranges between 100,000 and 200,000 articles per day [...] from around 75,000 news sources” [21] (p. 1). NewsReader uses an archive that “contains billions of articles, biographies, and reports” [37] (p. 1). SUMMA platform “[is] able to ingest 400 TV streams simultaneously” [42] (p. 6). Hence, the components of JKPs are designed considering their performance to minimise the processing and distribution times [12,15]. JKPs also integrate legacy components and facilitate interoperability with other systems and external services [16,24,35,37,41]. All these factors make the software architecture of JKPs complex and difficult to maintain without guidance.

## 5. Future Directions for Research on JKPs

### 5.1. Implications for Research

#### 5.1.1. Stakeholders

Studies on understanding how journalists embrace digital tools can aid in better adapting JKPs to the way journalists work. Such studies should consider the journalists' perceptions on using intelligent systems for creating news, how journalists process and use background information and the journalists' experiences working with AI, etc. Along these lines, related studies have been proposed, but not limited to, the journalists' usage of social media for gathering and verifying information [65,66] and the relation of the journalism practices and AI [67,68]. Similar user-oriented studies should be conducted on readers and younger and future generations of news consumers to identify what new forms of interaction and consumption are more appealing to them. These studies could consider, for example, the readers' perceptions of automated journalism [69,70] and young people's engagement with news recommendations [71].

#### 5.1.2. Information

To date, the knowledge extraction and recognition of entities from images and videos remain limited. Due to that, JKPs are not able to capture enough information from multimedia news. Promising directions for extracting knowledge from multimedia sources are multimodal machine learning approaches [72] that combine different types of data such as visual and text representations [73,74] and spoken language understanding tasks that analyse and detect audio speech [75]. Another limitation for knowledge extraction is the dark entities (i.e., those entities that do not exist yet in the knowledge base) [76,77]. Fresh stories about newer facts are the most attractive news, therefore, the chances of finding entity representations for those newer facts in knowledge bases are low. Therefore, research on knowledge extraction from multimedia news and dark entities can improve news representation in JKPs.

#### 5.1.3. Functionalities

Non-technical users find it difficult to perform complex searches in knowledge bases, archives and background information due to their lack of expertise. The usage of chatbots can aid user interaction using natural language [42,78]. Additional solutions that can support journalists' interaction with knowledge and information, and automate news production are text summarisation [61], automated reporting or story generation [79,80] and automatic data visualisation [81]. Augmented reality may also bring new possibilities for assisting the exploration of information using knowledge representations and LOD [82].

#### 5.1.4. Techniques

Due to the increase in misinformation and propaganda, it is crucial for journalists and readers to detect and distinguish trustworthy information from fake and biased news. Hence, research on JKPs should include automating the detection of fake news, political bias and rumours across social media platforms and news sources [58,83]. Techniques for such purposes can benefit from research on automating fact-checking [17,59], detecting derived or copied works [21], and media and audio forensics to identify manipulated or tempered multimedia files [84,85]. In addition, identifying misinformation items before they are stored in the knowledge base can improve the data quality of JKPs. Another promising direction is the inclusion of neural-symbolic AI [86] techniques as part of the different components of JKPs. Neural-symbolic AI combines neural networks with reasoning and logic. This can facilitate the inference and deductive reasoning over the data in the JKPs and reduce the computational cost of reasoning over knowledge graphs [87].

#### 5.1.5. Components

In addition to automatic techniques for verification and fact-checking, promising collaborative tools for news and social media verification that involve journalists and

readers [88] should be considered, for example, the tools developed in the ReVeal (<https://revealproject.eu> accessed on 15 May 2022), InVID (<https://www.invid-project.eu> accessed on 15 May 2022) [89] and WeVerify (<https://weverify.eu> accessed on 15 May 2022) [90] projects. Some of these tools such as WeVerify employ blockchain and knowledge graphs services for recording debunked claims and news. These collaborative repositories could be considered as additional information sources from which JKPs can obtain checked claims and provenance information but also contribute with verified information. Apart from this, the current JKPs are focused on in-house platforms that are typically accessed through a computer and oriented to print journalism. However, there is limited research on components that can facilitate access to the services offered by JKPs for mobile journalism [91] (i.e., journalism edited and published through smartphones and oriented towards audio-visual storytelling).

#### 5.1.6. Concerns

There are no gold standards or methodologies to evaluate JKPs. Accordingly, research needs to include the design and study of evaluation methods for JKPs. Moreover, readers and journalists may perceive results from JKPs as less transparent and difficult to understand [92] as they are driven by AI. To improve their perception of trustworthiness and transparency, research on JKPs should consider explainable AI methods [93].

### 5.2. Implications for Practice

#### 5.2.1. Stakeholders

To date, there have not been any studies on the implementation of JKPs in newsrooms. Such studies should evaluate the effectiveness, adoption and demand of JKPs. The experiences in implementing JKPs can help to draw a digitalisation path for newsrooms by providing best practices and identifying the main obstacles and solutions. This can support newsrooms with the definition of their roadmaps towards the adoption of JKPs, as it facilitates the identification of the most relevant aspects of JKPs and particular needs according to their current stage. Related studies have considered and provided guidelines for the utilisation of AI in news creation processes in a broader sense [55].

#### 5.2.2. Information

The literature is unclear on how JKPs should best represent events and there is no general agreement on what constitutes an event [21]. Events can range from fine-grained actions like a shot, injury or a handshake between two actors [15] to bigger and broader events like the Spanish Civil War and the COVID-19 pandemic [23] or events in between like a trial process. Therefore, research on JKPs needs to define and discuss how different types of events at different granularity can co-exist in a JKP and what conceptualisations of the event are useful for specific use cases.

#### 5.2.3. Functionalities

A better understanding of how to represent events and news items can bring new possibilities for JKPs, for example, on data analysis like measuring the popularity of people and companies [15], finding cause and effect relations [21], and identifying newsworthy events for specific audiences and particular user' interests [18,33,94].

#### 5.2.4. Techniques

One of the main limitations of the studied JKPs is the extraction of enough and precise information from text and multimedia to represent news stories in high detail [19,31]. For the knowledge graph-based JKPs we have considered in this paper, this means representing the content of text and multimedia as knowledge graphs. JKPs use relation extraction models to extract the textual relations between the entities in news text [15,62]. However, these models are in an early research stage and the extracted relations are basic

and limited for representing news [95]. Therefore, the functionalities that are based on these models must be considered for the longer term.

#### 5.2.5. Components

Current open-source large triple-stores are not scalable and their reasoning services are time-consuming and use too many computing resources. This limits the possibilities for JKPs to exploit reasoning capabilities and analyse large knowledge graphs. Hence, scalable triple-stores and mechanisms for better reasoning over large knowledge graphs can ease the incorporation of such solutions and bring new possibilities for JKPs. A promising approach is the inclusion of entity spaces [96]. These are vector spaces that represent the different entities of a knowledge graph and also capture their semantic information. They can be used to speed up processes that require complex graph explorations like inferring and disambiguating knowledge for unseen entities. Another promising approach for integrating and managing information from different types of databases is the usage of virtual knowledge graph [97]. Virtual knowledge graphs represent the schema of the different databases and provide mechanisms for querying the databases using SPARQL, hence, it integrates databases on the schema level and reduces data replication.

#### 5.2.6. Concerns

Only the most recent projects proposed systems to deal with big data [37,39,42]. Their architectures must also keep the machine learning models up-to-date and replace them for future best-of-breed, facilitate the schema evolution of knowledge bases and ease the expansion, distribution and independence of services [44]. Research on software reference architectures [98] for JKPs can assist in better designing and implementing them, as well as establishing a vocabulary and a framework to compare JKPs.

### 6. Limitations

This study only covers the English-language literature and is based on JKPs developed in Europe, Canada and USA. We have not identified any relevant JKPs in other geographical regions, but of course such JKPs may have been reported in languages other than English. The study is also influenced by the authors' involvement in the development of News Hunter. To reduce bias, we have not included our JKP during the meta-analysis process and we limited the News Hunter contribution to supporting and extending the findings. Additionally, the purpose of our analysis is to review the current state and future directions of the field, and not to evaluate the quality of the proposals.

### 7. Conclusions

This study has addressed which challenges and opportunities have motivated knowledge graph-based JKPs (RQ1), how knowledge graph-based JKPs are addressing these challenges and opportunities (RQ2), and the future directions of research on knowledge graph-based JKPs (RQ3). To our knowledge, no previous studies have identified and analysed JKPs as an emerging type of intelligent information system in this way. Although there are examples of such systems in the literature, to date, ours is the first clear definition and broad analysis of JKPs and their context.

In current newsroom workflows, metadata annotation is a manual, time-consuming and error-prone process. Newsrooms face difficulties to implement high-quality information systems. Journalists spend a lot of their time monitoring and filtering vast volumes of news, time that otherwise could be invested in creative tasks. These vast volumes of data often lack fine-grained annotation and are split into different repositories with different schemas. This limits the capacity of newsrooms to analyse and exploit their information resources, and share data with news consumers. To help with these processes, newsrooms use a large variety of services that are challenging to integrate and operate together, hindering their evolution towards digitally-integrated newsrooms. JKPs are a new type of intelligent information system that offer many opportunities for high-quality journalism in



newsrooms by combining AI, knowledge bases, LOD, NLP, ML and deep learning techniques. JKPs automate the metadata annotation and content enrichment with background information from external sources; monitor internal and word-wide news media output; facilitate event detection; support news creation and verification. They also facilitate the ingestion of vast amounts of data, and its storage, organisation and distribution. JKPs can provide newsrooms with a digitalisation path to reduce production costs and improve information quality while adapting the current workflows of newsrooms to new forms of journalism and readers' demands. We expect the next generation of JKPs to focus on enhancing journalism and providing unexpected news insights for journalists.

Many JKPs are big-data-oriented systems [15,21,22,35,42,44] that need a significant investment effort from newsrooms, making the adoption of JKPs challenging for small or local newsrooms. The adoption of JKP can yield many benefits, but newsrooms may perceive JKPs as an investment risk and look for alternative services. Thus, the formalisation of JKPs and the usage of open-source and out-of-the-box solutions, together with the popularisation of knowledge graphs will lower the adoption risk and increase the benefits. For small and local newsrooms, sharing JKPs can reduce the entrance barriers—a practice that is becoming more popular among digital-born news organisations and freelancers.

Section 5 has already proposed several paths for further research on JKPs. As an immediate continuation of this study, we are designing the software reference architecture for JKPs and developing tools to further study and enhance JKPs [44]. Through this work, we plan to define a reference model for JKPs that will allow their comparison and validation.

**Author Contributions:** M.G.O.: conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, visualization, writing—original draft. A.L.O.: conceptualization, resources, data curation, writing—review and editing, supervision, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Norwegian Research Council IKTPLUSS project 275872.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The study is influenced by the authors' involvement in the development of the News Hunter platform. To reduce bias, we have not included our JKP during the meta-analysis process and we limited the News Hunter contribution to supporting and extending the findings. Additionally, the purpose of our analysis is to review the current state and future directions of the field, and not to evaluate the quality of the proposals. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

JKP	Journalistic Knowledge Platform
AI	Artificial Intelligence
ML	Machine Learning
NLP	Natural Language Processing
LOD	Linked Open Data
RDF	Resource Description Framework

## Appendix A. Analysis Method

To synthesise data from the literature on the platforms that fit under our definition of JKP, we have used a qualitative meta-analysis approach [99,100]. We have searched the research literature and identified 28 papers describing 14 JKPs carried by distinct research groups located in different countries and in collaboration with a variety of news organisations and industry partners (Table 1 presents an overview of the selected JKPs and

papers). According to Maxwell [101], our sample represents an adequate variation in the phenomenon of interest. During the meta-analysis process, we focused on the last 10 years of advances and excluded our JKP from the process of extracting and coding data. After we synthesised the first conclusions, the excluded JKPs were added and analysed to support and expand our findings. This decision was taken to focus on the most recent advances and minimise the bias in the meta-analysis process inducted from our point of view.

From the selected literature we manually extracted 322 claims about the JKPs, i.e., statements that described the current state or expressed potential challenges or opportunities. Two independent expert coders (viz., the authors) conducted a purposive sampling [102,103] using the extracted claims that became marked up with 406 codes. We cleaned the generated codes with the support of NLP and natural language understanding techniques (implemented in python with support of Scikit-learn [104], NLTK [105], SpaCy [106] and other libraries) (i.e., Damerau-Levenshtein distance [107], word2vec [108] and Wordnet [109]). After cleaning and tidying up the initial codes, we iteratively classified the resulting codes into six top-level categories and 64 sub-categories (Figures 1 and 2 and Tables 2–5 shown the final top-level and sub-categories).

We used the top-level and sub-categories to re-code the 322 claims and we measured the final agreement using Gwet's  $AC_1$  [110] inter-rater reliability coefficient with nominal ratings. The  $AC_1$  coefficient for each category was 0.77 for Stakeholders, 0.65 for Components, 0.71 for Techniques, 0.71 for Aspects, 0.72 for Information types and 0.57 for Functionalities; with an average  $AC_1$  of 0.69 and a standard deviation of 0.063. Following the recommendations of Gwet [110] about Landis–Koch and Altman's benchmark scales, our  $AC_1$  express an acceptable agreement among coders. Finally, we agreed on the final codes for each claim and initiated an abductive process to understand and derive the current state and future directions of the research of JKPs.

## References

1. Newman, N.; Fletcher, R.; Schulz, A.; Andi, S.; Robertson, C.T.; Nielsen, R.K. *Reuters Institute Digital News Report 2021*; Technical Report; Reuters Institute for the Study of Journalism: Oxford, UK, 2021.
2. Simon, F.M.; Graves, L. *Pay Models for Online News in the US and Europe: 2019 Update*; Technical Report; Reuters Institute for the Study of Journalism: Oxford, UK, 2019.
3. Fletcher, R.; Nielsen, R.K. Paying for Online News. *Digit. J.* **2017**, *5*, 1173–1191. [CrossRef]
4. Newman, N.; Fletcher, R.; Kalogeropoulos, A.; Nielsen, R.K. *Reuters Institute Digital News Report 2019*; Technical Report; Reuters Institute for the Study of Journalism: Oxford, UK, 2019.
5. Newman, N.; Fletcher, R.; Schulz, A.; Andi, S.; Nielsen, R.K. *Reuters Institute Digital News Report 2020*; Technical Report; Reuters Institute for the Study of Journalism: Oxford, UK, 2020.
6. Toff, B.J.; Badrinathan, S.; Mont'Alverne, C.; Arguedas, A.R.; Fletcher, R.; Nielsen, R.K. *Overcoming Indifference: What Attitudes Towards News Tell Us about Building Trust*; Technical Report; Reuters Institute for the Study of Journalism: Oxford, UK, 2021.
7. Newman, N.; Fletcher, R. *Bias, Bullshit and Lies: Audience Perspectives on Low Trust in the Media*; Technical Report; SSRN: Rochester, NY, USA, 2017. [CrossRef]
8. Vázquez Herrero, J.; Direito-Rebollal, S.; Rodríguez, A.S.; García, X. *Journalistic Metamorphosis: Media Transformation in the Digital Age*; Springer: Cham, Switzerland, 2020. [CrossRef]
9. Beckett, C. *New Powers, New Responsibilities: A Global Survey of Journalism and Artificial Intelligence*; Technical Report; Polis, London School of Economics and Political Science: London, UK, 2019.
10. Lewis, S.C.; Westlund, O. Big Data and Journalism. *Digit. J.* **2015**, *3*, 447–466. [CrossRef]
11. Keefe, J.; Zhou, Y.; Merrill, J.B. The Present and Potential of AI in Journalism. Knight Foundation. 2021. Available online: <https://knightfoundation.org/articles/the-present-and-potential-of-ai-in-journalism/> (accessed on 22 February 2022).
12. Fernández, N.; Blázquez, J.M.; Fisteus, J.A.; Sánchez, L.; Sintek, M.; Bernardi, A.; Fuentes, M.; Marrara, A.; Ben-Asher, Z. NEWS: Bringing Semantic Web Technologies into News Agencies. In Proceedings of the Semantic Web—ISWC 2006, Athens, GA, USA, 5–9 November 2006; pp. 778–791. [CrossRef]
13. Maiden, N.; Zachos, K.; Brown, A.; Brock, G.; Nyre, L.; Nygård Tonheim, A.; Apsotolou, D.; Evans, J. Making the News: Digital Creativity Support for Journalists. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–11. [CrossRef]
14. Castells, P.; Perdrix, F.; Pulido, E.; Rico, M.; Benjamins, R.; Contreras, J.; Lorés, J. Neptuno: Semantic Web Technologies for a Digital Newspaper Archive. In Proceedings of the Semantic Web: Research and Applications, ESWS 2004, Heraklion, Crete, Greece, 10–12 May 2004. [CrossRef]

15. Rospocher, M.; van Erp, M.; Vossen, P.; Fokkens, A.; Aldabe, I.; Rigau, G.; Soroa, A.; Ploeger, T.; Bogaard, T. Building Event-Centric Knowledge Graphs from News. *J. Web Semant.* **2016**, *37–38*, 132–151. [[CrossRef](#)]
16. Raimond, Y.; Scott, T.; Oliver, S.; Sinclair, P.; Smethurst, M. Use of Semantic Web technologies on the BBC Web Sites. In *Linking Enterprise Data*; Springer: New York, NY, USA, 2010. [[CrossRef](#)]
17. Miranda, S.A.; Nogueira, D.; Mendes, A.; Vlachos, A.; Secker, A.; Garrett, R.; Mitchel, J.; Marinho, Z. Automated Fact Checking in the News Room. In Proceedings of the World Wide Web Conference, WWW '19, San Francisco, CA, USA, 13–17 May 2019; Association for Computing Machinery: New York, NY, USA; pp. 3579–3583. [[CrossRef](#)]
18. Kalfoglou, Y.; Domingue, J.; Motta, E.; Vargas-Vera, M.; Buckingham Shum, S. myPlanet: An ontology driven Web based personalised news service. In Proceedings of the International Joint Conference on Artificial Intelligence, Washington, DC, USA, 4–10 August 2001; Volume 2001, pp. 44–52.
19. Java, A.; Finin, T.; Nirenburg, S. SemNews: A Semantic News Framework. In Proceedings of the Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, Boston, MA, USA, 16–20 July 2006.
20. Borsje, J.; Levering, L.; Frasinca, F. Hermes: A Semantic Web-Based News Decision Support System. In Proceedings of the 2008 ACM Symposium on Applied Computing, SAC '08, Fortaleza, Brazil, 16–20 March 2008; Association for Computing Machinery: New York, NY, USA, 2008; pp. 2415–2420. [[CrossRef](#)]
21. Leban, G.; Fortuna, B.; Brank, J.; Grobelnik, M. Event Registry: Learning about World Events from News. In Proceedings of the 23rd International Conference on World Wide Web, WWW'14 Companion, Seoul, Korea, 7–11 April 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 107–110. [[CrossRef](#)]
22. Liu, X.; Nourbakhsh, A.; Li, Q.; Shah, S.; Martin, R.; Duprey, J. Reuters tracer: Toward automated news production using large scale social media data. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 1483–1493.
23. Rudnik, C.; Ehrhart, T.; Ferret, O.; Teyssou, D.; Troncy, R.; Tannier, X. Searching News Articles Using an Event Knowledge Graph Leveraged by Wikidata. In Proceedings of the Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, San Francisco, CA, USA, 13–17 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1232–1239. [[CrossRef](#)]
24. Ramagem, D.B.; Margerin, B.; Kendall, J. AnnoTerra: Building an integrated earth science resource using semantic Web technologies. *IEEE Intell. Syst.* **2004**, *19*, 48–57. [[CrossRef](#)]
25. Al-Moslimi, T.; Gallofré Ocaña, M.; Opdahl, A.L.; Tessem, B. Detecting Newsworthy Events in a Journalistic Platform. In Proceedings of the 3rd European Data and Computational Journalism Conference, Malaga, Spain, 1–2 July 2019; pp. 3–5.
26. Hogan, A.; Blomqvist, E.; Cochez, M.; D'Amato, C.; Melo, G.D.; Gutierrez, C.; Kirrane, S.; Gayo, J.E.L.; Navigli, R.; Neumaier, S.; et al. Knowledge Graphs. *ACM Comput. Surv.* **2021**, *54*, 1–257. [[CrossRef](#)]
27. Bizer, C.; Heath, T.; Berners-Lee, T. Linked data: The story so far. In *Semantic Services, Interoperability and Web Applications: Emerging Concepts*; IGI Global: Hershey, PA, USA, 2011; pp. 205–227.
28. Opdahl, A.L.; Al-Moslimi, T.; Dang-Nguyen, D.T.; Gallofré Ocaña, M.; Tessem, B.; Veres, C. Semantic Knowledge Graphs for the News: A Review. *Comput. Surv.* **2022**, to appear.
29. Berven, A.; Christensen, O.A.; Moldeklev, S.; Opdahl, A.L.; Villanger, K.J. A knowledge-graph platform for newsrooms. *Comput. Ind.* **2020**, *123*, 103321. [[CrossRef](#)]
30. Gallofré Ocaña, M.; Opdahl, A.L. Challenges and Opportunities for Journalistic Knowledge Platforms. In Proceedings of the CIKM 2020 Workshops, Galway, Ireland, 19–23 October 2020.
31. Domingue, J.; Motta, E. PlanetOnto: From news publishing to integrated knowledge management support. *IEEE Intell. Syst. Their Appl.* **2000**, *15*, 26–32. [[CrossRef](#)]
32. Frasinca, F.; Borsje, J.; Levering, L. A semantic web-based approach for building personalized news services. *Int. J. E-Bus. Res. (IJEBR)* **2009**, *5*, 35–53. [[CrossRef](#)]
33. Schouten, K.; Ruijgrok, P.; Borsje, J.; Frasinca, F.; Levering, L.; Hogenboom, F. A semantic web-based approach for personalizing news. In Proceedings of the 2010 ACM Symposium on Applied Computing—SAC '10, Sierre, Switzerland, 22–26 March 2010; ACM Press: Sierre, Switzerland, 2010; p. 854. [[CrossRef](#)]
34. Kobilarov, G.; Scott, T.; Raimond, Y.; Oliver, S.; Sizemore, C.; Smethurst, M.; Bizer, C.; Lee, R. Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections. In *The Semantic Web: Research and Applications*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5554. [[CrossRef](#)]
35. Fernández, N.; Fuentes, D.; Sánchez, L.; Fisteus, J.A. The NEWS ontology: Design and applications. *Expert Syst. Appl.* **2010**, *37*, 8694–8704. [[CrossRef](#)]
36. Kattenberg, M.; Beloki, Z.; Soroa, A.; Artola, X.; Fokkens, A.; Huygen, P.; Verstoep, K. Two architectures for parallel processing for huge amounts of text. In Proceedings of the Language Resources and Evaluation Conference (LREC). European Language Resources Association (ELRA), Portorož, Slovenia, 23–28 May 2016; pp. 4513–4519.
37. Vossen, P.; Agerri, R.; Aldabe, I.; Cybulska, A.; van Erp, M.; Fokkens, A.; Laparra, E.; Minard, A.L.; Aprosio, A.P.; Rigau, G.; et al. NewsReader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. *Spec. Issue Knowl.-Based Syst. Elsevier* **2016**, *110*, 60–85. [[CrossRef](#)]

38. Li, Q.; Shah, S.; Liu, X.; Nourbakhsh, A.; Fang, R. TweetSift: Tweet Topic Classification Based on Entity Knowledge Base and Topic Enhanced Word Embedding. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, Indianapolis, IN, USA, 24–28 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 2429–2432. [CrossRef]
39. Liu, X.; Li, Q.; Nourbakhsh, A.; Fang, R.; Thomas, M.; Anderson, K.; Kociuba, R.; Vedder, M.; Pomerville, S.; Wudali, R.; et al. Reuters Tracer: A Large Scale System of Detecting & Verifying Real-Time News Events from Twitter. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, Indianapolis, IN, USA, 24–28 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 207–216. [CrossRef]
40. Paikens, P.; Barzdins, G.; Mendes, A.; Ferreira, D.C.; Broscheit, S.; Almeida, M.S.; Miranda, S.; Nogueira, D.; Balage, P.; Martins, A.F. SUMMA at TAC Knowledge Base Population Task 2016. In Proceedings of the Ninth Text Analysis Conference (TAC), Gaithersburg, MA, USA, 14–15 November 2016.
41. Germann, U.; Liepins, R.; Gosko, D.; Barzdins, G. Integrating Multiple NLP Technologies into an Open-source Platform for Multilingual Media Monitoring. In Proceedings of the Workshop for NLP Open Source Software (NLP-OSS): Melbourne, Australia, 19–20 July 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 47–51. [CrossRef]
42. Germann, U.; Liepins, R.; Barzdins, G.; Gosko, D.; Miranda, S.; Nogueira, D. The SUMMA Platform: A Scalable Infrastructure for Multi-lingual Multi-media Monitoring. In Proceedings of the ACL 2018, System Demonstrations: Melbourne, Australia, 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 15–20 July 2018; pp. 99–104. [CrossRef]
43. Gallofré Ocaña, M.; Nyre, L.; Opdahl, A.L.; Tessem, B.; Trattner, C.; Veres, C. Towards a Big Data Platform for News Angles. In Proceedings of the 4th Norwegian Big Data Symposium (NOBIDS) 2018, Trondheim, Norway, 14 November 2018; pp. 17–29.
44. Gallofré Ocaña, M.; Opdahl, A.L. Developing a Software Reference Architecture for Journalistic Knowledge Platforms. In Proceedings of the ECSA2021 Companion Volume, Växjö, Sweden, 13–17 September 2021.
45. Berners-Lee, T.; Hendler, J.; Lassila, O. The semantic web. *Sci. Am.* **2001**, *284*, 34–43. [CrossRef]
46. Shadbolt, N.; Berners-Lee, T.; Hall, W. The Semantic Web Revisited. *IEEE Intell. Syst.* **2006**, *21*, 96–101. [CrossRef]
47. International Press Telecommunications Council. IPTC: Media Topics. 2022. Available online: <https://iptc.org/standards/media-topics/> (accessed on 22 February 2022).
48. Cyganiak, R.; Wood, D.; Lanthaler, M. RDF 1.1 Concepts and Abstract Syntax. 2014. Available online: <http://www.w3.org/TR/rdf11-concepts/> (accessed on 22 February 2022).
49. Troncy, R. Bringing the IPTC News Architecture into the Semantic Web. In Proceedings of the Semantic Web—ISWC 2008, Karlsruhe, Germany, 26–30 October 2008; Sheth, A., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T., Thirunarayan, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 483–498.
50. International Press Telecommunications Council. IPTC: NewsCodes. 2022. Available online: <https://iptc.org/standards/newscodes/> (accessed on 22 February 2022).
51. Miles, A.; Bechhofer, S. SKOS Simple Knowledge Organization System Namespace Document—HTML Variant. 2009. Available online: <http://www.w3.org/2004/02/skos/core.html> (accessed on 22 February 2022).
52. W3C OWL Working Group. OWL 2 Web Ontology Language Document Overview (Second Edition). 2012. Available online: <https://www.w3.org/TR/owl-overview/> (accessed on 22 February 2022).
53. International Press Telecommunications Council. IPTC: News Architecture. 2022. Available online: <https://iptc.org/standards/news-architecture/> (accessed on 22 February 2022).
54. Opdahl, A.L.; Tessem, B. Ontologies for finding journalistic angles. *Softw. Syst. Model.* **2020**, *20*, 71–87. [CrossRef]
55. Lopez, M.G.; Porlezza, C.; Cooper, G.; Makri, S.; MacFarlane, A.; Missaoui, S. A Question of Design: Strategies for Embedding AI-Driven Tools into Journalistic Work Routines. *Digit. J.* **2022**, *10*, 1–20. [CrossRef]
56. Gutierrez Lopez, M.; Makri, S.; MacFarlane, A.; Porlezza, C.; Cooper, G.; Missaoui, S. Making newsworthy news: The integral role of creativity and verification in the human information behavior that drives news story creation. *J. Assoc. Inf. Sci. Technol.* **2022**, *online version of record*.
57. Deuze, M. On creativity. *Journalism* **2019**, *20*, 130–134. [CrossRef]
58. Meel, P.; Vishwakarma, D.K. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Syst. Appl.* **2020**, *153*, 112986. [CrossRef]
59. Guo, Z.; Schlichtkrull, M.; Vlachos, A. A Survey on Automated Fact-Checking. *Trans. Assoc. Comput. Linguist.* **2022**, *10*, 178–206.
60. Diakopoulos, N. Computational News Discovery: Towards Design Considerations for Editorial Orientation Algorithms in Journalism. *Digit. J.* **2020**, *8*, 945–967. [CrossRef]
61. El-Kassas, W.S.; Salama, C.R.; Rafea, A.A.; Mohamed, H.K. Automatic text summarization: A comprehensive survey. *Expert Syst. Appl.* **2021**, *165*, 113679. [CrossRef]
62. Al-Moslmi, T.; Gallofré Ocaña, M. Lifting News into a Journalistic Knowledge Platform. In Proceedings of the CIKM 2020 Workshops, Galway, Ireland, 19–23 October 2020.
63. Garlan, D. Software Architecture. In *Encyclopedia of Software Engineering*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008. [CrossRef]
64. Gallofré Ocaña, M.; Al-Moslmi, T.; Opdahl, A.L. Data Privacy in Journalistic Knowledge Platforms. In Proceedings of the CIKM 2020 Workshops, Galway, Ireland, 19–23 October 2020.



65. Neuberger, C.; Nuernbergk, C.; Langenohl, S. Journalism as Multichannel Communication. *J. Stud.* **2019**, *20*, 1260–1280. [[CrossRef](#)]
66. Zhang, X.; Li, W. From Social Media with News: Journalists' Social Media Use for Sourcing and Verification. *J. Pract.* **2020**, *14*, 1193–1210. [[CrossRef](#)]
67. Stray, J. Making Artificial Intelligence Work for Investigative Journalism. *Digit. J.* **2019**, *7*, 1076–1097. [[CrossRef](#)]
68. Broussard, M.; Diakopoulos, N.; Guzman, A.L.; Abebe, R.; Dupagne, M.; Chuan, C.H. Artificial Intelligence and Journalism. *J. Mass Commun. Q.* **2019**, *96*, 673–695. [[CrossRef](#)]
69. Graefe, A.; Bohlken, N. Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News. *Media Commun.* **2020**, *8*, 50–59. [[CrossRef](#)]
70. Tandoc, E.C., Jr.; Yao, L.J.; Wu, S. Man vs. Machine? The Impact of Algorithm Authorship on News Credibility. *Digit. J.* **2020**, *8*, 548–562. [[CrossRef](#)]
71. Swart, J. Experiencing Algorithms: How Young People Understand, Feel About, and Engage with Algorithmic News Selection on Social Media. *Soc. Media Soc.* **2021**, *7*, 20563051211008828. [[CrossRef](#)]
72. Guo, W.; Wang, J.; Wang, S. Deep Multimodal Representation Learning: A Survey. *IEEE Access* **2019**, *7*, 63373–63394. [[CrossRef](#)]
73. Mogadala, A.; Kalimuthu, M.; Klakow, D. Trends in integration of vision and language research: A survey of tasks, datasets, and methods. *J. Artif. Intell. Res.* **2021**, *71*, 1183–1317. [[CrossRef](#)]
74. Chen, S.; Aguilar, G.; Neves, L.; Solorio, T. Can images help recognize entities? A study of the role of images for Multimodal NER. In Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021), Online, 11 November 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 87–96. [[CrossRef](#)]
75. Shon, S.; Pasad, A.; Wu, F.; Brusco, P.; Artzi, Y.; Livescu, K.; Han, K.J. SLUE: New Benchmark Tasks For Spoken Language Understanding Evaluation on Natural Speech. In Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 7927–7931. [[CrossRef](#)]
76. van Erp, M.; Ilievski, F.; Rospocher, M.; Vossen, P. Missing Mr. Brown and buying an Abraham Lincoln—Dark entities and DBpedia. In Proceedings of the Third NLP & DBpedia Workshop, Bethlehem, PA, USA, 11 October 2015; pp. 81–86.
77. Al-Moslmi, T.; Gallofré Ocaña, M.; Opdahl, A.L.; Veres, C. Named entity extraction for knowledge graphs: A literature overview. *IEEE Access* **2020**, *8*, 32862–32881. [[CrossRef](#)]
78. Luo, B.; Lau, R.Y.; Li, C.; Si, Y.W. A critical review of state-of-the-art chatbot designs and applications. *WIREs Data Min. Knowl. Discov.* **2022**, *12*, e1434. [[CrossRef](#)]
79. Miroshnichenko, A. AI to Bypass Creativity. Will Robots Replace Journalists? (The Answer Is “Yes”). *Information* **2018**, *9*, 183. [[CrossRef](#)]
80. Alhussain, A.I.; Azmi, A.M. Automatic Story Generation: A Survey of Approaches. *ACM Comput. Surv.* **2021**, *54*, 1–38. [[CrossRef](#)]
81. Zhu, S.; Sun, G.; Jiang, Q.; Zha, M.; Liang, R. A survey on automatic infographics and visualization recommendations. *Vis. Inform.* **2020**, *4*, 24–40. [[CrossRef](#)]
82. Lampropoulos, G.; Keramopoulos, E.; Diamantaras, K. Enhancing the functionality of augmented reality using deep learning, semantic web and knowledge graphs: A review. *Vis. Inform.* **2020**, *4*, 32–42. [[CrossRef](#)]
83. Zhou, X.; Zafarani, R. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Comput. Surv.* **2020**, *53*, 1–40. [[CrossRef](#)]
84. Pasquini, C.; Amerini, I.; Boato, G. Media forensics on social media platforms: A survey. *EURASIP J. Inf. Secur.* **2021**, *2021*, 1–19. [[CrossRef](#)]
85. Bhagtani, K.; Yadav, A.K.S.; Bartusiak, E.R.; Xiang, Z.; Shao, R.; Baireddy, S.; Delp, E.J. An Overview of Recent Work in Media Forensics: Methods and Threats. *arXiv* **2022**, arXiv:2204.12067.
86. Hitzler, P.; Bianchi, F.; Ebrahimi, M.; Sarker, M.K. Neural-symbolic integration and the Semantic Web. *Semant. Web* **2020**, *11*, 3–11. [[CrossRef](#)]
87. Hitzler, P.; Krotzsch, M.; Rudolph, S. *Foundations of Semantic Web Technologies*; CRC Press: Boca Raton, FL, USA, 2010. [[CrossRef](#)]
88. Thomson, T.; Angus, D.; Dootson, P.; Hurcombe, E.; Smith, A. Visual Mis/disinformation in Journalism and Public Communications: Current Verification Practices, Challenges, and Future Opportunities. *J. Pract.* **2020**, *16*, 1–25. [[CrossRef](#)]
89. Collyda, C.; Apostolidis, E.; Pournaras, A.; Markatopoulou, F.; Mezaris, V.; Patras, I. VideoAnalysis4ALL: An On-Line Tool for the Automatic Fragmentation and Concept-Based Annotation, and the Interactive Exploration of Videos. In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, ICMR '17, Bucharest, Romania, 6–9 June 2017; Association for Computing Machinery: New York, NY, USA, 2017. [[CrossRef](#)]
90. Marinova, Z.; Spangenberg, J.; Teyssou, D.; Papadopoulos, S.; Sarris, N.; Alaphilippe, A.; Bontcheva, K. Weverify: Wider and Enhanced Verification for You Project Overview and Tools. In Proceedings of the 2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW), London, UK, 6–10 July 2020; pp. 1–4. [[CrossRef](#)]
91. Salzmann, A.; Guribye, F.; Gynnild, A. “We in the Mojo Community”—Exploring a Global Network of Mobile Journalists. *J. Pract.* **2021**, *15*, 620–637. [[CrossRef](#)]
92. Shin, D. Why Does Explainability Matter in News Analytic Systems? Proposing Explainable Analytic Journalism. *J. Stud.* **2021**, *22*, 1047–1065. [[CrossRef](#)]
93. Kaur, D.; Uslu, S.; Rittichier, K.J.; Durrresi, A. Trustworthy Artificial Intelligence: A Review. *ACM Comput. Surv.* **2022**, *55*, 1–38. [[CrossRef](#)]



94. Motta, E.; Daga, E.; Opdahl, A.L.; Tessem, B. Analysis and Design of Computational News Angles. *IEEE Access* **2020**, *8*, 120613–120626. [[CrossRef](#)]
95. Yan, Y.; Sun, H.; Liu, J. A Review and Outlook for Relation Extraction. In Proceedings of the 5th International Conference on Computer Science and Application Engineering, CSAE 2021, Sanya, China, 19–21 October 2021; Association for Computing Machinery: New York, NY, USA, 2021. [[CrossRef](#)]
96. van Erp, M.; Groth, P. Towards Entity Spaces. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; European Language Resources Association: Marseille, France, 2020; pp. 2129–2137.
97. Xiao, G.; Ding, L.; Cogrel, B.; Calvanese, D. Virtual Knowledge Graphs: An Overview of Systems and Use Cases. *Data Intell.* **2019**, *1*, 201–223. [[CrossRef](#)]
98. Martínez-Fernández, S.; Ayala, C.P.; Franch, X.; Marques, H.M. Benefits and drawbacks of software reference architectures: A case study. *Inf. Softw. Technol.* **2017**, *88*, 37–52. [[CrossRef](#)]
99. Eisenhardt, K.M. Building Theories from Case Study Research. *Acad. Manag. Rev.* **1989**, *14*, 532–550. [[CrossRef](#)]
100. Hoon, C. Meta-Synthesis of Qualitative Case Studies: An Approach to Theory Building. *Organ. Res. Methods* **2013**, *16*, 522–556. [[CrossRef](#)]
101. Maxwell, J.A. *A Realist Approach for Qualitative Research*; Sage: Newcastle upon Tyne, UK, 2012.
102. Corbin, J.; Strauss, A. Grounded theory research: Procedures, canons, and evaluative criteria. *Qual. Sociol.* **1990**, *13*, 2–21. [[CrossRef](#)]
103. Corbin, J.; Strauss, A. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*; Sage Publications: Newcastle upon Tyne, UK, 1998.
104. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
105. Bird, S.; Loper, E.; Klein, E. *Natural Language Processing with Python*; O'Reilly Media, Inc.: Newton, MA, USA, 2009.
106. Honnibal, M.; Montani, I.; van Landeghem, S.; Boyd, A. *spaCy: Industrial-Strength Natural Language Processing in Python*; Zenodo: Geneva, Switzerland, 2020. [[CrossRef](#)]
107. Damerau, F.J. A Technique for Computer Detection and Correction of Spelling Errors. *Commun. ACM* **1964**, *7*, 171–176. [[CrossRef](#)]
108. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of the Advances in Neural Information Processing Systems 26, NIPS 2013, Barcelona, Spain, 11–19 December 2013; Neural Information Processing Systems Foundation: San Diego, CA, USA.
109. Miller, G.A. WordNet: A Lexical Database for English. *Commun. ACM* **1995**, *38*, 39–41. [[CrossRef](#)]
110. Gwet, K.L. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters*; Advanced Analytics, LLC: Oxford, MS, USA, 2014.