EMBEDDIA

Cross-Lingual Embeddings for Less-Represented Languages in European News Media

Research and Innovation Action

Call: H2020-ICT-2018-1

Call topic: ICT-29-2018 A multilingual Next generation Internet

Project start: 1 January 2019 Project duration: 36 months

D6.5: Refined analysis of news media partners' needs and challenges (T6.1)

Executive summary

The purpose of this report is to present an updated overview of the user needs in the news media industry for the EMBEDDIA project. In this refined analysis with respect to artificial intelligence (AI) tools in newsrooms we partly build upon “D6.3: User needs and challenges for news media industry”. The previous report derived from an initial project workshop in March 2019 where user needs were explored, applying action-research paired with user-oriented design as the methodology. There were three key recommendations about what EMBEDDIA should focus on: comment management, detection of interesting news, and personalized news generation. Here, this preliminary guidance is expanded based on what has turned out to be important, preferable, and technically feasible. We continued using the same methodology but are here focusing on creating three user stories based on interviews with journalists and managers from the media companies to facilitate a shared understanding of the collaboration. These user stories paint a picture of affordances and what journalists could achieve with EMBEDDIA newsroom technology tools.

Partner in charge: UH

Project co-funded by the European Commission within Horizon 2020 Dissemination Level		
PU	Public	PU
PP	Restricted to other programme participants (including the Commission Services)	-
RE	Restricted to a group specified by the Consortium (including the Commission Services)	-
CO	Confidential, only for members of the Consortium (including the Commission Services)	-

This project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement No 825153

Deliverable information

Document administrative information	
Project acronym:	**EMBEDDIA**
Project number:	**825153**
Deliverable number:	**D6.5**
Deliverable full title:	**Refined analysis of news media partners' needs and challenges**
Deliverable short title:	**Refined analysis of user needs**
Document identifier:	**EMBEDDIA-D65-RefinedAnalysisOfUserNeeds-T61-submitted**
Lead partner short name:	**UH**
Report version:	**submitted**
Report submission date:	**30/06/2020**
Dissemination level:	**PU**
Nature:	**R = Report**
Lead author(s):	**Carl-Gustav Lindén (UH)**
Co-author(s):	**Silver Traat (TEXTA), Martin Kjellman (STT), Martin Marzidovšek (JSI), Senja Pollak (JSI), Matthew Purver (QMUL), Ivar Krustok (ExM), Marko Pranjić (TRI)**
Status:	**_ draft, _ final, x submitted**

The EMBEDDIA Consortium partner responsible for this deliverable has addressed all comments received. Changes to this document are detailed in the change log table below.

Change log

Date	Version number	Author/Editor	Summary of changes made
15/04/2020	v1.0	Carl-Gustav Lindén (UH)	Draft report template.
23/04/2020	v1.1	Matthew Purver (QMUL), Senja Pollak (JSI), Marko Pranjić (TRI), Ivar Krustok (ExM)	User stories for ExM and 24sata.
23/04/2020	v1.2	Martin Kjellman (STT)	User stories for STT and methodology.
18/05/2020	v1.3	Carl-Gustav Lindén (UH)	Updated version.
19/05/2020	v.1.4	Martin Kjellman (STT)	Comments on updated version.
22/05/2020	v1.5	Leo Leppänen (UH)	Comments on updated version.
22/05/2020	v1.6	Hannu Toivonen (UH)	Comments on updated version.
26/05/2020	v1.7	Silver Traat (TEXTA)	Contributing 2.5 User story driven development.
27/05/2020	v1.8	Carl-Gustav Lindén (UH)	Report sent to reviewers.
28/05/2020	v1.9	Marko Milosavljević (UL)	Reviewer comments.
05/06/2020	v1.10	Matthew Purver (QMUL)	Reviewer comments.
08/06/2020	v1.11	Carl-Gustav Lindén (UH)	Incorporated reviewers' comments.
09/06/2020	v1.12	Nada Lavrač (JSI)	Report quality checked and finalised.
24/06/2020	v1.13	Martin Kjellman (STT), Matthew Purver (QMUL), Senja Pollak (JSI)	Additional background on user stories.
24/06/2020	final	Carl-Gustav Lindén (UH)	Final version.
30/06/2020	submitted	Tina Anžič (JSI)	Report submitted.

Table of contents

List of abbreviations

AI	Artificial Intelligence
API	Appliction Programming Interface
CMS	Content Management System
DoA	Description of Action
EC	European Commission
ExM	Ekspress Meedia
GUI	Graphical User Interface
NLG	Natural Language Generation
NLP	Natural Language Processing
REST	Representational State Transfer
STT	Finnish News Agency
TRI	Trikoder
WP	Work Package

1 Introduction

The purpose of this report is to present an updated overview of the user needs in the news media industry for the EMBEDDIA project. In this refined analysis with respect to artificial intelligence (AI) tools in newsrooms, we partly build upon “D6.3: User needs and challenges for news media industry”. The previous report derived from a project workshop where user needs were explored, applying action-research paired with user-oriented design as the methodology. There were three key recommendations about what EMBEDDIA should focus on: comment management, detection of interesting news, and personalized news generation. Here, this preliminary guidance is expanded based on what has turned out to be important, preferable, and technically feasible. We continue with a similar methodology but are here focusing on user stories based on interviews with journalists and managers from the media companies.

1.1 Main purpose of EMBEDDIA

While advanced automated language technology tools and resources exist for a few dominant languages (English, French, German), many of Europe's language communities - and the news media industry that serves them - lack appropriate tools for multilingual internet and text-based industry development. EMBEDDIA aims to solve some of these limitations and develop methods and tools, an EMBEDDIA Media Assistant, to allow to transfer resources and tools from technologically well-supported languages to others. The techniques and technologies developed in the project will enable journalists, editors and researchers to search, link and monitor news reports and editorial content; analyse and react to public user comments; produce content semi-automatically, and to do all this across European languages.The aim is to provide tools for smaller and medium-sized media companies and newsrooms in under-resourced languages that might otherwise lag behind.

Three small and medium-sized European media companies are part of the project. These are Ekspress Grupp from Estonia (ExM), Trikoder (TRI) from Croatia, and the Finnish News Agency (STT). They have different profiles. ExM is a leading media group in the Baltic states focusing on publishing, printing services, and online media content production. Trikoder is part of Styria Group, one of the leading media groups in Austria, Croatia, and Slovenia with a broad offering of services ranging from newspapers to books and radio stations. STT is the only news agency in Finland. The publicly available EMBEDDIA Media Asistant will be developed for them and other media companies by the Estonian language technology company Texta.

1.2 Development and adoption of new technology in media and journalism

Investing in new technology and reforming newsrooms around artificial intelligence is costly and strategically complicated (Agrawal et al., 2018). It is less about the costs of investment in software and more of choosing the right editorial strategy and making sure that the newsroom is equipped with the right skills for radical transformation. This becomes a problem specifically for small and medium sized media companies. Due to several reasons, such as the downfall of advertising revenues and changing news consumption habits, the media industry is struggling not just to reinvent the business model but for survival (Lehtisaari et al, 2012). The lack of an innovation culture does not help either. Instead, imitation and copying are core elements of media innovation (Boczkowski, 2005). For instance, a study of US metro papers shows that companies are mostly rearranging existing business models rather than experimenting with radical innovations, while examples of new projects breaking the barriers of legacy media are rare (Villi et al., 2019). Especially newspapers tend rather to reproduce the concepts of past

successes, focusing on incremental changes, rigorous brand alignment and top down-monitored, commercially steered activities (Järventie-Thesleff, Moisander, and Villi 2014). Media executives often hesitate to make bold, high-risk moves because the landscape keeps changing so fast and the risks with costly investments is high (McDowell 2011). The brakes on change are also cultural (Ess, 2014). Going from products to services, from hardware to software, and from audience to users and consumers includes changing mindsets, many times unlearning the trade and its institutional truths. The problems with adaption has become acute during the Covid-19 crisis which have struck news media hard with rapidly decreasing advertising revenues and income in general: this have reduced their capabilities to invest in new technologies even faster and makes the EMBEDDIA project even more important, as provider of insights and tools for the European media industry.

1.3 Artificial intelligence, news media and journalism

The emergence of a new technology tends to be viewed as an incremental process, with the new inevitably replacing the old technology. However, examples from history show an interactive process wherein each shapes the other (Chadwick 2013, 25-26). Artificial intelligence provides a technological challenge that is different from previous transformations as it is not just about new hardware, say radio or television, but a whole new human-machine system. Artificial intelligence (AI), automation, and machine learning are transforming journalism around the world. AI is already affecting the news value chain, from news gathering to content processing, creation and distribution (Marconi & Siegman, 2017; Weghe, 2018). New technology is for instance used to verify input sent by users, such as pictures or videos, to filter out unwanted user comments, to check facts, to personalise and enhance the reading experience through content recommendation, to automatically generate texts and video, or to use advance audience metrics based on user behaviour to analyse performance of content. For instance, the New York Times is using predictive analytics tools to gain a competitive edge. This includes so called funnel analysis to see how people become subscribers, and how to influence more to do so. They also use natural language processing to understand what content topics generate reader engagement, helping marketing teams to select articles to promote (Burns, 2015). At BBC an AI tool called Juicer aggregates news and extracts content. It watches 850 RSS feeds globally to take and tag articles at scale. Reuters tracer is a large scale system for detecting and verifying real-time news events from Twitter (Liu et al., 2016).

However, journalists struggle to understand what AI means and usually focus on industry announcements. As an example, the most common news source for AI in British news media has been the Silicon Valley entrepreneur Elon Musk (Brennen, Howard, & Nielsen, 2018). Among journalists and media managers, there is frustration about the lack of clarity around key definitions in AI, for instance the connection between intelligence and artificial. The hype around AI does not make the conversation easier: is this material substance or is this just the new shiny trend of the moment (Perretti, 2019; Milosavljević & Vobič, 2019)? Media companies, constrained by financial and human resources as well as skills, need to think hard about where to invest and innovate. Therefore, the adoption of AI applications needs to be preceded by an editorial strategy with a clear vision and sense of which elements media companies should explore and what they should leave aside (Beckett, 2019; Jääskeläinen & Olij, 2019). This confusion is certainly not confined solely to journalism. Despite 70 years of discussion, there is still disagreement on how to exactly define artificial intelligence. In this report, we rely on two similar definitions that underline the human side of AI, a ”human-centered future” (Diakopoulos, 2019) which come closer to the concept of augmented intelligence first promoted by Doug Engelbart (1962).

AI can be dfined as« the theory and development of computer systems able to perform tasks normally requiring human intelligence (Hansen, Roca-Sales, Keegan, & King, 2017).

Research on artificial intelligence in journalism is still a small field with a limited number of articles published. The most recent contribution is a collection of expert opinions edited by Seth Lewis (2019). There is more research on specific features of AI in journalism, for instance texts generated by Natural Language Generation systems (Diakopoulos, 2019; Dörr, 2016; Leppänen et al., 2017; Linden et al., 2019; Sirén-Heikel et al., 2019), recommender systems (Beam, 2014; Lianget al., 2006), bias detection in texts (Ali et al., 2010), content moderation with machine learning (Jiang & Han, 2019; Roberts, 2019), and story-finding and news worthiness or augmented creative writing (Carlson, 2018; Huovelin et al., 2013; Magnusson et al., 2016; Plattner, et al., 2017; Zachos et al., 2018) but almost nothing on a crucial aspect, the data that drives AI applications in newsrooms. Skills for working with numbers, large and small data sets, public records, and data visualizations are essential in news organizations today (Boyles & Meyer, 2017; Rogers et al., 2017; Weber et al., 2018). The literature on newsrooms strategies for data is scarce, for instance when it comes to news text generation, where the unhindered access to structural data is a crucial element of operations (Karlsson, 2019). Some strategies for exploring, evaluating and utilising data have been described in the literature (Linden et al., 2019; Magnusson et al., 2016), for instance with software that enables journalists to encode news events and stories directly as data (Caswell et al., 2015; Caswell & Dörr, 2019).

Certain specific features of newswork provide a challenge. Looking at the potential for AI in news reporting, Jonathan Stray notes that investigative journalism is a hard problem for AI, due to a number of reasons. One important aspect is the unique character of each investigation and the lack of structured data, which makes it hard to create computer models (Stray, 2019).

To sum up and prioritise this overview of the state of the art in the news media industry, we can see that most media outlets, which use new digital tools including AI systems/tools/assistants, use them for the following tasks:

- News search or aggregation: this helps contributors to find interesting topics or angles.
- Personalisation: tools used for user personalisation of content (articles, newsletters) including geo-targeting.
- Comment moderation: the mentioned tools are used for comment moderation or analysis and data extraction.
- General purposes: help in producing stories, either the tools write the stories independently or they propose headlines, topics or images.
- Other tasks: new digital tools include also data visualization, audience engagement, and transcription.

During the Covid-19 crisis, there has been a growth in AI applications for newsrooms that have helped journalists to report on the pandemic more effectively: Media companies such as Times/Times Sunday or Swedish Aftonbladet use text automation to outpace competitors with its covid-19 news coverage; BBC launched a pop-up Corona Bot service to answer questions on Covid-19 related changes; Bloomberg uses a news detection AI tool to sift through sources and catch breaking news. These are just a few examples showing the rapid development of reporting tools.

2 User stories methodology

Insights in the previous report D6.3 that this report builds upon derived from a project workshop where user needs were explored, applying action-research paired with user-oriented design as the methodology. There were three key recommendations about what EMBEDDIA should focus on: comment management, detection of interesting news, and personalized news generation. Here, this

preliminary guidance is expanded based on – what during almost a year of discussions and experiments – has turned out to be important, preferable, and technically feasible. However, we continue with the same methodology which is essentialy about bridging communication problems – what do people need and how are these needs expressed and adressed? One one side there are users, in our case journalists with an organisational and skills perspective; on the other side there are researchers and developers with their own preferences, who might lack domain knowledge and struggle to understand user needs. Here we are focusing on user stories that build upon interviews with journalists and managers from the media companies. The methodology for this report is still based on action-research theory (Defrijna et al., 2008) combined with user-oriented design (Veryzer et al., 2005). We utilised both methods as they provide an orientation that fosters a deeper understanding of user needs and what constitutes value in practical terms for journalists. In this report we more precisely focus on user-stories which bridges both approaches (Cohn, 2004).

2.1 Action research

Action research offers a framework for research collaborations between scholars and practitioners with the clear intention to induce change and improvement of practice (Grubenmann, 2016). Its goal is to serve as a tool for ”solving problems experienced by people in their professional community” (Appelgren & Nygren, 2014). This research framework has been labelled a “remarkably inclusive methodology” (Cunningham, 2014, 3) and a “collaborative approach” (Stringer, 1996, 15). We argue that action research—and participatory action research, in particular—is less a methodology than an orientation or stance toward the research process and the participants (Cammarota & Fine, 2008). The main idea is that only an active participant can understand the dynamics of a specific human system, in this case the complexity of media innovations. Without trying to change it, the system remains invisible for the passive observer (Schein, 1987). In line with Dupagne and Chuan (2019), we argue that journalism and news media would be best served by ”being proactive toward AI development”. In the Tallinn EMBEDDIA workshop[1], we presented a conceptual and methodological framework for interactively and iteratively clarifying and designing software according to the needs of the EMBEDDIA media partners. Following Defrijna et al. (2008), in the EMBEDDIA project we apply cycles of planning, acting, observing, and reflection to the development of new AI tools for journalists.

The main reason behind this iterative approach is that a problem-driven and solutions-oriented research agenda helps us to orientate towards the specific conditions of the media partners involved in EMBEDDIA and the most relevant issues. This approach also puts researchers in a special position, for some previously unknown. The scientific bottom-up and participatory approach is motivated by a genuine interest in collaborating with people in the conduct of their work, as well as the future of their organizations and their communities. One could say that researchers involved are taking the role of a “friendly outsider” (Greenwood & Levin, 2006, 124-128). Furthermore, Grubenmann (2016) have defined three principles of action research:

- The intention to change. This solutions-driven approach distinguishes action research from other types of participatory research.
- Participatory and inclusive research. The goal is to maximize the usefulness of research outcomes for a particular community. Outcomes are local and problem specific. However, in the EMBEDDIA project we go further and believe that research outcomes will be generalizable and/or replicable.

[1] March 12-13, 2019. In the event media managers, journalists, and developers from media partners worked with researchers.

- Developmental research process. Outcomes are answers to community-relevant questions created in negotiation with each other. Knowledge is uncertain and ambiguous, answers tentative and open to modification. This is especially true in the field of media innovations, where new technology is developed and deployed rapidly in a trial and error fashion with uncertain financial outcomes while editorial resources and skills are scarce.

One specific challenge in action research is the terminology used to discuss key concepts and challenges. People from different disciplines or communities of experts have their own vocabularies for framing and presenting their thoughts. Metaphors play a central role in decision making. How they are actually interpreted and used can have a significant impact on decision outcomes, e.g., (Bosmajian, 1992; Hibbitts, 1994; Napoli, 1999; Wagemans & Witschge, 2019). Insofar as language does not just describe reality but constitutes reality, metaphors do not just describe but shape and create organizational strategy (Inayatullah et al., 2016).

2.2 User oriented design

The need for tools that are easy to understand and use is not new but the digital transformation of work and communication accentuates conceptual problems. People responsible for the creation of our digital products rarely take into account the users' goals, needs, or motivations. This results in products that lack a coherent user experience (Cooper et al., 2003). User oriented design is a powerful tool for answering the most important questions that crop up during the definition and design of a digital product. The phrase user-oriented design is used to underline the need for user design consideration in the context of digital product development that transforms a bundle of technology with the ability to provide functionality into a "product" that people desire to interact with and from which they derive benefits (Veryzer & Borja de Mozota, 2005, 128). That approach fosters a deeper appreciation of user needs and what delivers value to customers through evaluating problems and proposing solutions in the context of relevant user experience dimensions. The optimal result is products that provide maximum benefit to users. Table 1 contains a comparison of user-oriented and traditional approaches to designing products.

Table 1: Traditional and new approach to designing products (Veryzer & Borja de Mozota, 2005, 132; Vredenberg et al., 2001, 2).

Traditional Approach	**User-Oriented Design**
Technology driven	User driven
Component focus	Solutions focus
Limited multidisciplinary cooperation	Multidisciplinary team work
Focus on internals architecture	Focus on externals design
No specialization in user experience	Specialization in user experience
Some competitive focus	Focus on competition
Development prior to user validation	Develop only user validated designs
Product defect view of quality	User view of quality
Limited focus on user measurement	Prime focus on user measurement
Focus on current customers	Focus on current and future customers

In the case of EMBEDDIA, we have discussed a number of important issues and broadly follow a list of questions.

- Who are the users?
- What are users trying to accomplish?
- How do users think about what they're trying to accomplish?
- What kind of experiences do users find appealing and rewarding?
- How should the product behave?
- What forms should the product take?
- How will users interact with the product?

To facilitate an ongoing discussion of these issues, we created a user project committe that convened twice during the first half of 2020. There, a smaller group was able to discuss the user needs of journalists at the media partners in more detail.

2.3 Methodology for user stories

User stories describe functionality that will be valuable to users or buyers of software. They usually contain three aspects (Cohn, 2004, 4):

- A written description of the story used for planning and as a reminder.
- Conversations about the story that serve to flesh out the details of the story.
- Tests that convey and document details and that can be used to determine when a story is complete.

In software development, the user stories concept is a way of representing requirements for a system through brief texts that cover the three basic elements of a requirement: who is the system for, what does it need to do, and why is it important? (Lucassen et al., 2016a). In his seminal book on user stories, Cohn (2004, 3) stated, "software requirements is a communication problem" and proposed user stories as a means for customers/users and developers to handle the unpredictable nature of software development projects. During the last decade, user stories grew more popular and the concept remains a staple of requirements engineering (Kassab, 2015). Among the different variations of user stories, Lucassen et al. (2016a, 383) identify three aspects that they have in common: "(1) a short piece of text describing and representing the user story, (2) conversations between stakeholders to exchange perspectives on the user story, and (3) acceptance criteria". The user stories written for this report mainly represent the first and second of these aspects -- they are short textual representations derived at through conversations between different partners in the EMBEDDIA project; media representatives, researchers and developers. The aim of the EMBEDDIA project is not to develop commercially viable applications. Therefore, the planning for the final stage, acceptance tests and release of product, is not included in these user stories. The textual representations included in this report are also meant to spark further conversations between the stakeholders and guide the formulation of more specific acceptance criteria.

Practitioners who work with user stories tend to (a) perceive them as effective for reaching mutual understanding and making sure that the correct software is being developed and (b) find it beneficial to use a template (Lucassen et al., 2016b). The user stories written for this report use the template popularised by Cohn: "As a <type of user>, I want to <some goal>, so that <some reason>". In order to guarantee the usefulness of our user stories, we follow the Quality User Story framework developed by Lucassen et al. (2016a). This gives us two collections of quality criteria, one for the individual user story and one for a set of user stories. On the individual level, the user stories need to be:

- well-formed (includes the basic elements of a requirement, e.g., who the user is)
- atomic (describes only one feature)
- minimal (nothing more than the basic elements is included)
- conceptually sound (the basic elements are ones that go together)
- problem-oriented (does not suggest solutions)

- unambiguous (internally and in relationship to other stories)
- full sentence
- estimatable (the amount of work needed can be comprehended)

A set of user stories should be:

- unique and conflict-free
- uniform (stories follow the same format)
- independent (to the extent that it is possible a user story should not build directly upon another)
- complete (together the stories form a fully functioning application)

At this stage of the EMBEDDIA project, the last criterion on each level (estimatable and complete) proved most difficult to meet, suggesting that closer cooperation between the different stakeholders might be beneficial as the project progresses.
In this report, most of the focus is on how news automation is viewed at STT. The user stories for Ekspress Meedia and 24sata, part of the Styria Media Group, have a shorter description and explanation of the methodology for deriving the stories. The reason is that we wanted to test a more compherensive application of action research and user-oriented design, which was made possible with a research assistant situated in the STT newsroom.

2.3.1 Methodology for the STT user story

The following methodological section describes how user stories were built from interviews followed by a thematic content analysis. For the purpose of further exploring the needs and challenges of the media partners presented in the first user needs report (D6.3), five in-depth interviews were conducted with STT employees by an EMBEDDIA research assistant. For preparation of the interviews, newsroom practices and automation project meetings were observed during five weeks prior to the interviews.
Participants included newsroom staff as well as news managers. All participants had at least some experience with newsroom automation, either from previous projects, through their everyday work, or both. The sample was constructed so as to explore user needs on both newsroom and organisation level. The previous user needs report concluded that STT's customers, i.e. local and regional Finnish newspapers, are highly relevant end users. For this reason, STT's role as a service provider in the Finnish news ecosystem had to be taken into consideration.

In order to avoid any terminological confusion with regards to automation, respondents were asked to prepare for the interviews by reading summaries of the technology being developed in the EMBEDDIA project. Beforehand, respondents also received six background questions, the answers of which were used to adjust the interview guide. In addition to providing information about previous experience with newsroom automation, participants were asked to describe a potential automation use case, which they elaborated on during the interview through an interactive process.

The interview guide was constructed through an iterative process. First, a set of questions was created based on the results from the previous user needs report. Secondly, the questions were informed by topics that had proven useful in our other research on newsroom automation projects (Lindén et al., 2019).

Four of the interviews were conducted face-to-face. Due to coronavirus restrictions, the final interview was conducted using the video service Zoom. The language was Finnish, the mother-tongue of all the respondents. Since a Finnish translation was lacking for some of the key concepts, questions were

asked in English to avoid misunderstanding. Further, interviews were preceded by a brief discussion on terminology where respondents were provided with a list of key terms with translations. The interviews were recorded and lasted for 30 minutes to an hour. In the transcription process, the interviews were split up into meaning units and translated to English. In a second iteration, these units were transformed into themes to reflect what type of user need was being expressed.

2.3.2 Methodology for user stories at Ekspress Meedia and Trikoder/24sata

At Ekspress Meedia, a researcher from the EMBEDDIA project interviewed one manager twice online to create a user story on the needs for semi-automated keyword selection. At 24sata a meeting with journalists was arranged where two managers also participated.

In this section, we will present three user stories, which have a future perspective with imagined newsroom affordances (Nagy & Neff, 2015). The first short one (Table 2) focuses on one journalist at STT but the results also contains a broader overview of different user needs that were identified during observation and interviews. The two second ones (Table 3 and Table 4) focus on macro and micro perspectives of user stories in the STT newsroom. The user stories for Ekspress media (Table 5) and 24sata (Table 6) are based on the interviews and meetings as mentioned above.

3 User stories

3.1 User story: STT text generation

Table 2: User story: STT text generation.

As Head of news at the Finnish news agency Suomen tietotoimisto (STT), Virpi's job is to coordinate the news production. This includes deciding what topics to focus on, distributing assignments and making sure that STT's customers - regional newspapers that rely on STT for their supply of domestic and international news coverage - know what is being produced. On an average day, the newsroom produces 250-300 news items, and as head of news, Virpi must guarantee that the customers get accurate and swift reports. STT maintains close cooperation with other news agencies in the Nordic countries and around the world. Therefore, all content from Finland has to be translated into Swedish and English, which is now done automatically. *In recent years, news reporting has become increasingly data-driven. For these purposes, Virpi can usually rely on a small group of journalists who excel at finding and analysing relevant datasets, such as statistics from the Eurostat database. These tasks are time-consuming however, and due to diminishing resources, it is not possible to have anyone focus solely on this. Even during normal circumstances, the manpower Virpi is able to allocate towards data-driven reporting is limited, and ever since the coronavirus outbreak in early 2020, almost all resources are needed for more urgent tasks.* *Using EMBEDDIA technology, the job of analysing the datasets has become much more efficient as will be detailed in user stories below. The data still needs to be located and evaluated by journalists first, but the process of analysing potential news value and leads in the data is now handled using automation through which reporters quickly get an overview in short text format. This is a big time-saver that is particularly valuable in high-stress situations. Thanks to multilingual text generation Virpi can make sure that whenever new data comes in, the initial reports reach the customers - local and regional Finnish newspapers quickly without compromising the accuracy of the content. The EMBEDDIA text generator simultaneously produces the same texts in Finnish, Swedish and English*

that are also shared with the international partner network of news agencies, mainly in the Nordics. Thanks to automated text production, Virpi is now able to assign data-driven reporting to journalists who are not trained in statistics or computer science. When the EMBEDDIA text generator supplies them with natural language texts based on already analysed data, the journalists are able to use their own strengths: prioritising newsworthiness and writing articles in a way that the audience can understand.

3.1.1 Multilingual text generation

In the following part, we extend the discussion on user needs in the case of multilingual text generation. First, we note that the term "end user" still is somewhat unclear, since the term could refer to a newsroom employee as well as an audience member (see D6.3 for discussion). Which term is most accurate is in part dependent on how the automatically generated texts are used. Are they automatically published and thus immediately available to the reader, or are they used as raw materials for journalists and editors? In the previous user needs report, consensus leaned towards the latter. This view was confirmed in interviews with STT employees. Therefore, in this revised report, the term end user also refers to journalists and editors.

Based on the interviews, a second important distinction needs to be made and it is related to STT's role in the Finnish news ecosystem. STT does not only supply their customers (local and regional Finnish newspapers) with content. The news agency also aims to support their customers with services that support the reporting customers themselves are doing. Therefore, the end user can be either a journalist or editor in STT's own newsroom or the equivalent in the customer's newsroom.

The most important finding also relates to STT's role as a service provider in the Finnish news ecosystem. While text generation, and especially its multilingual features, is relevant for STT's own newsroom, the news agency's top priority is to enable their customers' data-driven reporting. Such an approach entails providing customers with relevant data and the means to understand and process the data for news reporting. Automated text generation based on data represents one possible solution to this need.

A second important implication of this service approach is that modification for different needs becomes highly relevant; the customers need to be able to choose what aspects of the data they want texts to cover. What aspect is relevant depends on which region of the country they cover, meaning that in this case modification is mainly of geographical nature. For example, public health service data from the municipalities covered by Newspaper A might not be interesting to Newspaper B and vice versa. However, the ability for customers themselves to control the modification process is considered important, since they possess the local expertise valued by their own audiences.

Regarding the functionality of the system, the key findings relate to journalists' skill sets and work practices. The participants all agree that most journalists - in STT's own newsroom as well as in the customers' newsrooms - lack the technical skills needed to efficiently use a complex system. This issue is exacerbated by a general lack of resources. Most newsrooms are stretched very thin due to financial struggles, meaning that for the majority of journalists, finding time to learn new skills poses a challenge. This makes simplicity a key requirement for any system being developed. The lack of resources also means that most news organisations lack the flexibility needed for adopting new work practices and tools. Another requirement, therefore, is that a new system is broadly applicable rather than specialised. Unifying the demands on simplicity and broad applicability is an important issue for the EMBEDDIA developers to address.

Finally, the main user need expressed vis-à-vis multilingual text generation concerns data handling and analysis. These skills are in short supply in both STT's and the customers' newsrooms. Hence, the respondents felt that the most significant added value provided by automated text generation is the ability to analyse data. As one respondent puts it when talking about reporting on coronavirus data:
"For the most part we are writers, specialised in the arts or social sciences. We understand society and how things relate to each other, but when it comes down to crunching numbers, that's not necessarily the easiest thing for everyone. In a sense that supports the notion of some automation providing the results that you can then go ahead and make conclusions based upon."

For a text generation system then the most important requirement is that it allows reporters to make sense of data by providing texts that do not have to be elaborate but clear and thus easy to understand.

3.1.2 Macro perspective user stories

As a news agency, STT has dual responsibilities: On the one hand, the company produces news in its own newsroom. On the other, it provides its customers (local and regional newspapers) with infrastructure for their own reporting. In the following two sections the macro perspective user stories deal with user needs related to STT's role as a provider of infrastructure, whereas the micro perspective user stories (Section 3.1.3) focus on the more specific needs of STT's own newsroom. Each row tells a specific user story. Each column represents one of the three basic elements of a requirement: who is the system for, what does it need to do, and why is it important? (Lucassen et al., 2016a).

Table 3: Macro perspective user stories.

AS A/AN	I WANT TO...	SO THAT...
customer	be able to choose particular data to report on	my reporting is relevant to the region in which members of my audience live
producer	distribute relevant data to my customers in a format that is useful to them	I provide some added value to their reporting
producer	help my customers make sense of the data	they can focus on what to report and how to present content to their audience
producer	let my customers access automatically generated text on-demand	they get to choose how they approach the data
customer	get automatically generated text that help me make sense of the data	I can use these texts as a starting point for more in-depth reporting
customer	locate and make sense of data that supports the story I am currently working on	I get context and added value for my article
producer	get all automatically produced texts in Swedish and English	I can share them with non-Finnish customers and news agencies

3.1.3 Micro perspective user stories

The micro perspective user stories highlight the kind of functionality prioritised by STT's own newsroom staff.

Table 4: Micro perspective user stories.

AS A/AN	I WANT TO...	SO THAT...
reporter	know that the texts are factually correct	my focus can be on presenting information in a way that is meaningful for my audience
reporter	get texts that are uncomplicated and easy to read	I can be sure I understand what I am reporting
reporter	be able to choose what aspects of the data on which to get text	I am in charge of determining newsworthiness
head of news	automate the production of news texts that always follow the same patterns	the newsroom can cover more of these stories or/and focus on topics that can't be automated
head of news	be able to capture recurring events such as traffic accidents as data	the newsroom can automate the production of news texts on these events
reporter	automatically generate headlines that build on previous headlines on the same topic	I save time and the headline is a logical continuation of past ones
sports reporter	get a text that gives me an overview of what happened in last night's NHL games	I can easily identify relevant information
sports reporter	know which Finnish players played last night and how they did	I can easily evaluate what is relevant to a Finnish audience
sports reporter	order background information on specific players or teams either as data or text	I can include relevant context in my article

3.2 User story: Ekspress Meedia semi-automated keyword selection

The user story was developed in collaboration with partner ExM. It focuses on linking the articles with keywords, which helps journalists looking for similar articles, editors when linking related articles, and readers when browsing the portal. Semi-automated procedure of keyword assignment, which is currently done manually, would save time to journalists and making tagging more consistent.

Table 5: User story: Ekspress Meedia semi-automated keyword selection.

Mari is a journalist at Ekspress Meedia (ExM), and she writes about 10 news articles a day. Once an article is written, she needs to define around five free keywords that best characterize the article. These keywords correspond to the main topics of the article, to help finding the articles of interest; ideally, the assigned keywords should be helpful for journalists looking for similar articles, for editors when linking related articles, and for readers when browsing the portal. *Until now, the job of keyword assignment at ExM had to be performed manually. This is a demanding task, though: the Estonian tagset at ExM has about 66,000 possible keywords. At least it was cut down recently — until just a few months ago there were more than 210,000 tags to choose from — but cutting them down to keywords that have been actually used was a difficult and time consuming job that doesn't happen often. So given the time pressure, Mari often has to choose keywords without paying much attention. She's sure that she doesn't choose the best keyword combination, and often prefers to assign keywords without checking whether they actually are in the ExM tagset.* *With the new EMBEDDIA techniques for automated keyword extraction and selection, though, Mari's job is much easier. The process is now semi-automated: candidate keywords are detected automatically and keywords already in the ExM tagset are prioritised, to make it easy for Mari to keep the tagset consistent. Keywords not in the tagset may also be suggested, subject to Mari's confirmation, and the tags will be processed by an editor in a later phase. Keywords which refer to named entities (names of persons, organizations, places) are also included. Mari now just has to make the final choice, selecting a few best keywords from the list of ten proposed top-ranked keywords for a given article. The selected keywords will be highlighted in the article where they first appear, with named entities highlighted in a different colour to make them easy to spot. Clicking on a keyword takes you to a page with all the articles with this tag.*

3.3 User story: 24sata comment moderation

Our third user story was developed in consultation with partner TRI. It focuses on the user-generated content that must be handled by publishers, in the form of comments posted by readers under online news articles. The comments section is a key factor in attracting readers, and provides a valuable arena for free speech and public access. However, they bring legal and ethical concerns for publishers, and therefore the main concern of our partners is the ability to filter and moderate the comments - see full story in Table 6.

Table 6: User story: 24sata comment moderation.

Branko works as a moderator at 24sata, the largest-circulation daily newspaper in Croatia. 24sata reaches about 2 million readers daily, and many of them post comments on its online articles: on an average day, about 8,000 comments come in, spread over several hundred articles. Unfortunately, many comments (usually between 5% and 10%) need to be blocked to prevent them appearing online: they might be offensive, dangerous or legally compromising. This is Branko's job. *Until now, the task of comment filtering and moderation had to be performed almost entirely manually. This is time-consuming and skilled work: the newspaper has a complex moderation policy, as comments may need blocking for a variety of reasons. Some are irrelevant spam or advertising, some contain disinformation, some are threatening or hateful, some obscene or illegal, some written in foreign languages … so filtering through them all and making consistent decisions is difficult, especially at peak times when over 1,000 per hour may be coming in. Branko uses a system which flags comments that*

match a list of blacklisted keywords, but this isn't very accurate and is hard to keep up to date as new topics get discussed. With the current COVID-19 crisis, for example, new kinds of spam, fake stories and ethnically-targeted hate speech emerge very fast, and the word lists can't keep up. That means Branko largely has to rely on fast reading and experience.

The new EMBEDDIA tools for automated comment moderation have made Branko's job much easier. Comments are filtered in real time, automatically detecting those which are most likely to need blocking, ranking them by severity, and labelling them as to which part of the 24sata policy they seem to break. The final decision is left to Branko, but now he can easily prioritise the worst cases first, and make sure they don't appear on the site, without having to read through all the others. He can then check less severe cases, and can leave unproblematic comments where the classifier is very confident for a less busy time. Branko's final decisions are then stored and fed back to the system, so that it learns over time to improve, and to adapt to new vocabulary as new topics and stories develop.

3.4 User story driven development

Based on the user stories described before, clear end user needs were discovered. Users with comments need automated comment filtering. Users writing articles need tools to generate keywords and topics automatically. Journalists who write news will benefit from tools that generate brief stories automatically. It also became evident that most newsrooms already have their own content management systems (CMS) and it is unlikely for them to drop them and adopt new tools. Therefore, we chose an API-driven approach. All separate services have REST APIs, they are wrapped together and dockerized. This makes integrating and using them rather painless. At this moment, Ekspress Media is already using one of the services in production to predict if comments should be moderated or not. For those users who would rather build their own models, we have configurated our Media Assistant so that it enables data browsing, building classification models and implementing them over REST API. Finally, for demo purposes we use the same APIs to show how everything works in real life through a GUI.

4 Associated outputs

We also publised a paper on the Data journalism as a service (majority of work was performed before official start of the project, but was published in 2020). For more details, see the following publication, which is also attached to this deliverable as an appendix.

Citation	Status	Appendix
Appelgren, E. & Lindén, C.G (2020). Data journalism as a service: Digital nativedata journalism expertise and product development. Media and Communication 8(2), 62–72.	Published	Appendix A

5 Conclusions and further work

The purpose of this report is to derive and prioritise user needs of the news media industry for the EMBEDDIA project. We analyse innovation challenges in news media and map the current state-of-the art in media technology with regards to artificial intelligence tools in newsrooms. Our previous user needs report presented the results from a project workshop where user needs were explored, applying action-research paired with user-oriented design as the methodology. The results from roundtable

discussions in the Tallinn workshop provided initial guidance for the development work in the project. In this revised follow-up report we focused on user stories, which turned out to be a valuable user-design tool that connected practioners, researchers and developers efficiently around shared insights, meanings and goals. The user stories presented here have been developed and tested on news managers involved in the project. However, this part would maybe not have been so successful without the social interaction that was a prerequisite for mutual understanding and shared language emerging from the workshop in Tallinn. In that sense, the methodology works progressively.

In the case of user stories, the methology was applied more extensively at the news agency STT with a research assistant situated in the newsroom for several months observing activities and building questions for semi-structured interviews upon these.

In the case of multilingual text generation, ”end user” has two different meanings since STT is a service provider to its media customers in the Finnish news ecosystem. The ”end user” can either a journalist or editor in STT’s own newsroom or the equivalent in the customer’s newsroom.

The key findings are:

- The most significant added value provided by automated text generation is the ability to analyse data, a system that allows reporters to make sense of data by providing texts that do not have to be elaborate but clear and thus easy to understand. While text generation, and especially its multilingual features, is relevant for STT’s own newsroom, the news agency’s top priority is to enable their customers’ data-driven reporting.
- Another important implication of this service approach is that modification for different needs becomes highly relevant; the customers need to be able to choose what aspects of the data they want texts to cover.
- Most journalists in STT’s own newsroom as well as in the customers’ newsrooms lack the technical skills needed to efficiently use a complex system. Therefore, a new system has to be broadly applicable rather than specialised.

Additional added value from this report comes from the methodological insight that action research and user-oriented design could be applied in a new fashion, as user stories. Having a research assistant situated in the newsroom with access to editors and journalists provided the project with some valuable insights: observations of newsroom work transformed into a questionnaire that contained issues that otherwise would have been invisible to researchers. Our methodological experience from both the workshop in Tallinn and the preparation for the revised report also inspired us to prepare a special issue of Nordicom Review, ”Action Research for Media Development: Intersections and boundaries of social change, innovation, and entrepreneurship”, that will be published in 2021. We also initiated a Nordic network of action research in media development including academics, media managers and journalists that will be expanded on an Euopean level in 2021.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction machines: The simple economics of artificial intelligence. Boston, Massachusetts: Harvard Business Press.

Ali, O., Flaounas, I., De Bie, T., Mosdell, N., Lewis, J., & Cristianini, N. (2010). Automating news content analysis: An application to gender bias and readability. Proceedings of the First Workshop on Applications of Pattern Analysis, 36-43.

Appelgren, E., & Nygren, G. (2014). Data journalism in Sweden: Introducing new methods and genres of journalism into “old” organizations. Digital Journalism, 2(3), 394-405.
Beam, M. A. (2014). Automating the news: How personalized news recommender system design choices impact news reception. Communication Research, 41(8), 1019-1041.
Beckett, C. (2019). New powers, new responsibilities: A global survey of journalism
and artificial intelligence. London: The London School of Economics and Political Science. Retrieved from https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/ (05.01.2020)
Boczkowski, P. J. (2005). Digitizing the news: Innovation in online newspapers. Cambridge, MA: MIT Press.
Bosmajian, H. A., & Bosmajian, H. A. (1992). Metaphor and reason in judicial opinions. Carbondale, Il: Southern Illinois University Press
Boyles, J. L., & Meyer, E. (2017). Newsrooms accommodate data-based news work. Newspaper Research Journal, 38(4), 428-438.
Brennen, J. S., Howard, P. N., & Nielsen, R. K. (2018). An industry-led debate: How UK media cover artificial intelligence. Reuters Institute for the Study of Journalism Fact Sheet.
Burns, E. (2015). How the New York Times uses predictive analytics algorithms. Available: http://searchbusinessanalytics.techtarget.com/feature/How-The-New-York-Times-uses-predictive-analytics-algorithms [29.12.2015]
Cammarota, J., & Fine, M. (Eds.). (2010). Revolutionizing education: Youth participatory action research in motion. New York: Routledge.
Carlson, M. (2018). Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. New Media & Society, 20(5), 1755-1772.
Caswell, D., & Dörr, K. (2019). Automating complex news stories by capturing news events as data. Journalism Practice, 13(8), 951-955.
Caswell, D., Russell, F., & Adair, B. (2015). Editorial aspects of reporting into structured narratives. Proceedings of the 2015 Computation Journalism Symposium.
Chadwick, A. (2013). The hybrid media system: Politics and power. Oxford: Oxford University Press.
Cohn, M. (2004). User stories applied: For agile software development. Boston: Addison-Wesley Professional.
Cunningham, J. (2014). Academic discourse. In D. Coghland, & M. Brydon-Miller (Eds.), The SAGE Encyclopedia of Action Research (pp. 1-3). Los Angeles, CA: SAGE Publications.
Cooper, A., Reimann, R., & Dubberly, H. (2003). About face 2.0: The essentials of interaction design Indianapolis, In: John Wiley & Sons, Inc.
Defrijna, S., Mathijsb, E., Gulinckc, H., & Lauwersa, L. (2008). Facilitating and evaluating farmer innovations towards more sustainable energy and material flows: Case-study in Flanders. Empowerment of the Rural Actors: A Renewal of Farming Systems Perspectives: 8th European IFSA Symposium, Clermont-Ferrand, France, 6-10 July 2008, 765-773.
Diakopoulos, N. (2019). Automating the news: How algorithms are rewriting the media. Cambridge, Massachusetts: Harvard University Press.
Dupagne, M., & Chuan, C. (2019). Concluding comments: A few more points about artificial intelligence. Journalism & Mass Communication Quarterly, 1-23.
Dörr, K. N. (2016). Mapping the field of algorithmic journalism. Digital Journalism, 4(6), 700-722.
Engelbart, D. C. (1962). Augmenting human intellect: A conceptual framework. Menlo Park, CA: Stanford Research Institute.
Ess, C. M. (2014). Editor's introduction: Innovations in the newsroom – and beyond. Journal of Media Innovations, 1(2), 1-9.
Greenwood, D. J., & Levin, M. (2006). Introduction to action research: Social research for social change. Thousand Oaks, CA: SAGE Publications.
Grubenmann, S. (2016). Action research. Digital Journalism, 4(1), 160-176.

Hansen, M., Roca-Sales, M., Keegan, J. M., & King, G. (2017). Artificial intelligence: Practice and implications for journalism. New York: Tow Center for Digital Journalism.
Hibbitts, B. J. (1994). Making sense of metaphors: Visuality, aurality, and the reconfiguration of American legal discourse. Cardozo L.Rev., 16, 229.
Huovelin, J., Gross, O., Solin, O., Linden, K., Maisala, S. P. T., Oittinen, T., Toivonen, H., Niemi, J., Silfverberg, M. (2013). Software newsroom–an approach to automation of news search and editing. Journal of Print Media Technology Research. 2(3), 141-1556.
Inayatullah, S., Izgarjan, A., Kuusi, O., & Minkkinen, M. (2016). Metaphors in futures research. Futures. 84, 109-114.
Jiang, L., & Han, E. H. (2019). Modbot: Automatic comments moderation. Proceedings of the Computation + Journalism Symposium
Järventie-Thesleff, R., Moisander, J., & Villi, M. (2014). The strategic challenge of continuous change in multi-platform media organizations—A strategy-as-practice perspective. International Journal on Media Management, 16(3-4), 123-138.
Jääskeläinen, A., & Olij, M. (2019). The next newsroom: Unlocking the power of AI for public service journalism. Geneva: EBU.
Karlsson, S. (2019). Structured data, not internet scraping, results in trustworthy robot-produced journalism. Retrieved from https://www.inma.org/blogs/big-data-for-news-publishers/post.cfm/structured-data-not-internet-scraping-results-in-trustworthy-robot-produced-journalism [26.08.2019]
Kassab, M. (2015, August). The changing landscape of requirements engineering practices over the past decade. In 2015 IEEE Fifth International Workshop on Empirical Requirements Engineering (EmpiRE) (pp. 1-8). IEEE.
Lehtisaari, K., Karppinen, K., Harjuniemi, T., Grönlund, M., Lindén, C., Nieminen, H., & Viljakainen, A. (2012). Media convergence and business models: Responses of Finnish daily newspapers. (No. 4/2012). Helsinki: Communication Research Centre CRC, University of Helsinki.
Leppänen, L., Munezero, M., Sirén-Heikel, S., Granroth-Wilding, M., & Toivonen, H. (2017). Finding and expressing news from structured data. Proceedings of the 21st International Academic Mindtrek Conference, 174-183.
Liang, T., Lai, H., & Ku, Y. (2006). Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings. Journal of Management Information Systems, 23(3), 45-70.
Linden, T. C., Tuulonen, H. E., Bäck, A., Diakopoulos, N., Granroth-Wilding, M., Haapanen, L., Leppänen, L., Melin, M., Moring, T., Munezero, M. D. (2019). News automation: The rewards, risks and realities of 'machine journalism'. Frankfurt. WAN-IFRA.
Lewis, S. C. (2019). Artificial intelligence and journalism. Journalism & Mass Communication Quarterly, 1-23.
Liu, X., Li, Q., Nourbakhsh, A., Fang, R., Thomas, M., Anderson, K., Kociuba, R., Vedder, M., Pomerville, S., Wudali, R. (2016). Reuters tracer: A large scale system of detecting & verifying real-time news events from Twitter. Proceedings of the 25th ACM International Conference on Information and Knowledge Management, 207-216.
Lucassen, G., Dalpiaz, F., van der Werf, J. M. E., & Brinkkemper, S. (2016a). Improving agile requirements: the quality user story framework and tool. Requirements Engineering, 21(3), 383-403.
Lucassen, G., Dalpiaz, F., van der Werf, J. M. E., & Brinkkemper, S. (2016b). The use and effectiveness of user stories in practice. In International working conference on requirements engineering: Foundation for software quality (pp. 205-222). Springer, Cham.
Magnusson, M., Finnäs, J., & Wallentin, L. (2016). Finding the news lead in the data haystack: Automated local data journalism using crime data. Computation + Journalism Symposium.
Marconi, F., & Siegman, A. (2017). The future of augmented journalism: A guide for newsrooms in the age of smart machines. New York: AP Insights.

McDowell, W. S. (2011). The brand management crisis facing the business of journalism. The International Journal on Media Management, 13(1), 37-51.
Milosavljević, M., & Vobič, I. (2019). 'Our task is to demystify fears': Analysing newsroom management of automation in journalism. Journalism. 1-19.
Nagy, P., & Neff, G. (2015). Imagined affordance: Reconstructing a keyword for communication theory. Social Media Society, DOI: 10.1177/2056305115603385.
Napoli, P. M. (1999). The marketplace of ideas metaphor in communications regulation. Journal of Communication, 49(4), 151-169.
Perretti, M. (2019, May 21). What journalists talk about when they talk about AI. Retrieved from https://blogs.lse.ac.uk/polis/2019/05/21/what-journalists-talk-about-when-they-talk-about-ai/ [18.06.2019]
Sirén-Heikel, S., Leppänen, L., Lindén, C., & Bäck, A. (2019). Unboxing news automation. Nordic Journal of Media Studies. 1(1), 47-66.
Stringer, E. T. (Ed.). (1996). Action research. A handbook for practitioners. Thousand Oaks, CA: SAGE Publications, Inc.
Roberts, S. T. (2019). Behind the screen: Content moderation in the shadows of social media. New Haven and London: Yale University Press.
Rogers, S., Schwabish, J., & Bowers, D. (2017). Data journalism in 2017: The current state and challenges facing the field today. Retrieved from https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf [16.08.2019]
Schein, E. H. (1987). The clinical perspective in fieldwork. Newbury Park, CA: Sage Publications, Inc.
Stray, J. (2019). Making artificial intelligence work for investigative journalism. Digital Journalism, 1- 22.
Veryzer, R. W., & Borja de Mozota, B. (2005). The impact of user-oriented design on new product development: An examination of fundamental relationships. Journal of Product Innovation Management, 22(2), 128-143.
Villi, M., Grönlund, M., Linden, C., Lehtisaari, K., Mierzejewska, B., Picard, R. G., & Röpnack, A. (2019). "They're a little bit squeezed in the middle": Strategic challenges for innovation in US metropolitan newspaper organisations. Journal of Media Business Studies, 1-18.
Vredenberg, K., Isensee, S., & Righi, C. (2001). User-centered design: An integrated approach. Upper Saddle River, NJ: Prentice Hall PTR.
Wagemans, A., & Witschge, T. (2019). Examining innovation as process: Action research in journalism studies. Convergence, 25(2), 209-224.
Weber, W., Engebretsen, M., & Kennedy, H. (2018). Data stories: Rethinking journalistic storytelling in the context of data journalism. SComS Studies in Communication Sciences, 2018 (1), 191-206.
Weghe, T. V. D. (2018). 10 things about AI every newsroom should know. Retrieved from https://medium.com/jsk-class-of-2019/10-things-about-ai-every-newsroom-should-know-19745dac8ad7 [09.08.2019]
Zachos, K., Apostolou, D., Paraskevopoulos, F., Ientsek, S., Maiden, N., Brown, A., & Mentzas, G. (2018). Creative information exploration in journalism. 9th International Conference on Information, Intelligence, Systems and Applications (IISA), 1-7.

Appendix A: Data journalism as a service: Digital native data journalism expertise and product development

COGITATIO

Media and Communication (ISSN: 2183–2439)
2020, Volume 8, Issue 2, Pages 62–72
DOI: 10.17645/mac.v8i2.2757

Article

Data Journalism as a Service: Digital Native Data Journalism Expertise and Product Development

Ester Appelgren [1,*] and Carl-Gustav Lindén [2]

[1] Department of Journalism, Södertörn University, 141 89 Huddinge, Sweden; E-Mail: ester.appelgren@sh.se
[2] Swedish School of Social Science, University of Helsinki, 00014 Helsinki, Finland; E-Mail: carl-gustav.linden@helsinki.fi

* Corresponding author

Submitted: 2 January 2019 | Accepted: 7 February 2020 | Published: 16 April 2020

Abstract
The combined set of skills needed for producing data journalism (e.g., investigative journalism methods, programming, knowledge in statistics, data management, statistical reporting, and design) challenges the understanding of what competences a journalist needs and the boundaries for the tasks journalists perform. Scholars denote external actors with these types of knowledge as interlopers or actors at the periphery of journalism. In this study, we follow two Swedish digital native data journalism start-ups operating in the Nordics from when they were founded in 2012 to 2019. Although the start-ups have been successful in news journalism over the years and acted as drivers for change in Nordic news innovation, they also have a presence in sectors other than journalism. This qualitative case study, which is based on interviews over time with the start-up founders and a qualitative analysis of blog posts written by the employees at the two start-ups, tells a story of journalists working at the periphery of legacy media, at least temporarily forced to leave journalism behind yet successfully using journalistic thinking outside of journalistic contexts.

Keywords
boundary work; data journalism; digital native; Finland; journalism; peripheral actors; Sweden

Issue
This article is part of the issue "Digital Native News Media: Trends and Challenges" edited by Ramón Salaverría (University of Navarra, Spain).

1. Introduction

Embedded in the terms 'digital native' or 'online native' is a precondition that there are (now) news sites without an analogue heritage. Harlow and Salaverría (2016) suggest that digital native news websites are not only innovative start-ups associated with new and alternative ways of producing and presenting journalism with differing degrees of political activism but also funded, or rather underfunded, by new business models. Furthermore, these initiatives are not limited to producing journalism, but also engage in the organization of cultural events, thus expanding on the business models for news production (García-Avilés, Carvajal-Prieto, Arias, & De Lara-González, 2019). According to Harlow and Salaverría (2016), digital native outlets tend to describe their use of technology as different from mainstream media, for example, using data journalism techniques, such as creating databases and infographics and making documents publicly available. These techniques may not be different from what legacy media use, but they are nevertheless associated with being innovative. Developing skills in technology and self-promotion are important for entrepreneurial journalists (Cohen, 2012), a group that includes data journalists. Data journalists, in turn, are often regarded as forerunners of the journalism of the future (Knight, 2015). Loosen, Reimer, and De Silva-Schmidt (2017) argue that data journalism is commonly carried out by cross-disciplinary teams that have divided the labor into data analysis, visualization and writing. When professionals and non-professionals produce news together in this manner, they are engaging in what Belair-Gagnon

and Holton (2018) define as boundary maintenance (news production) around a boundary object (news). Researchers have described the professionals traditionally outside of journalism who are interacting at this boundary as interlopers or actors at the periphery of journalism (e.g., Belair-Gagnon & Holton, 2018; Cheruiyot, Baack, & Ferrer-Conill, 2019; Eldridge, 2018). According to Carlson (2016), these boundaries provide shared ways of understanding news work, and the metajournalistic discourse that arises from these collaborations may challenge how news is produced and consumed.

Previous research on news sites born online have focused on stand-alone sites with a direct audience (e.g., Bruno & Kleis Nielsen, 2012). In this study, we aim to explore how digital news initiatives that act as an intermediary between data and legacy media could also be part of the digital native ecosystem. Such news start-ups publish content and journalistic products, but not with the intention to attract an audience. Rather, the intention is similar to that of the larger non-profit news organizations around the world, such as ProPublica or the International Consortium of Investigative Journalists, which is to attract new collaborations with the media and through these collaborations republish the news to a wider audience (Konieczna & Powers, 2017).

This exploratory study revolves around a case featuring two such digital native news start-ups: Journalism ++ Stockholm and Journalism Robotics/Newsworthy. These companies share founders and employees and have overlapping aims of harnessing data at the periphery of traditional journalism. While Journalism ++ Stockholm mainly produces traditional data journalism on commission for legacy media, Journalism Robotics/Newsworthy focuses on automating parts of the data journalistic working process, i.e., the collection and refinement of public data, and provides analyses of data as short reports for both journalistic and non-journalistic purposes.

Hermida and Young (2019) assert that a key for future research is to gauge who matters in journalism based on qualities other than traditional ones (p. 99). Thus, the narrow focus on two individuals in this case study can be motivated by their central status as journalism innovators in Sweden and the Nordic countries. Furthermore, in his seminal work on news automation, Diakopoulos (2019) frames these two start-ups as groundbreaking in an international comparison as well.

Much of recent scholarship on innovation in journalism revolves around boundary work, where new actors enter journalism as peripheral actors (Powers & Zambrano, 2016; Usher, 2017) or engage with journalists as in the case of the entanglements of civic tech and data journalism (Cheruiyot et al., 2019). This qualitative case study is based on an analysis of blog posts and interviews over time with the start-up founders. The study contributes to the growing body of literature on peripheral actors in journalism by exploring digital natives who do not have a direct audience and apply a reversed perspective, thus presenting a story of journalists who work at the periphery of legacy media and use journalistic thinking outside of journalistic contexts.

In the following sections, we will briefly review the literature on digital news start-ups and relate data journalism to a subsection of the discussion on peripheral actors in journalism. We will then present the methods, material and motivating questions for the study, followed by our results. We conclude with a discussion and conclusions.

2. Literature Review

In this case study of two news start-ups, we focus on individuals who drive innovation and entrepreneurship in journalism from the periphery. Thus, the selected literature review is based on studies of entrepreneurial journalism, data journalism and peripheral actors, as well as factors of success in digital news start-ups.

2.1. The Rise and Success of Digital News Start-Ups

The literature on digital news start-ups revolves around start-ups producing journalism for a direct audience. In a report on such news start-ups in western Europe, Bruno and Kleis Nielsen (2012) found that online start-ups begin as journalistic and are driven by both professional and commercial ambitions. They tend to diversify their revenues by experimenting with activities that traditionally have not been associated with journalism, such as content syndication, e-commerce, advertorials, consultancy work, events planning or reader donations (p. 96). The scholarly discussion links this development of experimentation with new technologies and challenging of boundaries for what is considered journalism to the downsizing of the media industry. In this context, downsizing has had a positive effect by encouraging journalists to start their own businesses and propose their own journalistic innovations (Cohen, 2015; Raviola, 2019; Tenor, 2019). Cook and Sirkkunen (2015) divided digital native start-up sites into two categories: storytelling and service-orientation. Storytelling sites provide news to a direct audience, and service-oriented sites consist of online media actors that focus on elements of journalism rather than content creation. The revenue streams for service-oriented sites are not based on advertising models but rather on a wider strategy to sell products as part of a complex relationship with mainstream media (Cook & Sirkkunen, 2015, p. 73). Thus, previously unthreatened boundaries between journalistic and business-oriented functions are dissolving, often rhetorically motivated by survival and an industry crisis (Coddington, 2015). However, the change may not be that drastic. Naldi and Picard (2012) argue that entrepreneurs bring "formal myopia" into their new enterprise, which means their previous experiences and perceptions affect their assessments. Formal myopia, according to Naldi and Picard (2012, p. 77), influences initial thinking and choices of news start-up founders, thus limiting innovation and creating additional obsta-

cles to their success. However, Wagemans, Witschge, and Deuze (2016) argue that a strong belief in the core values of journalism, such as maintaining a professional identity, may not necessarily obstruct the development of new forms of entrepreneurial journalism; in fact, it can be the main selling point.

Some digital natives have been associated with counter information as a way to promote social change in societies facing democratic challenges and low confidence in the media (Harlow & Salaverría, 2016). There are also digital natives that focus on investigative journalism without distinct political or activist aims (Nee, 2013). Furthermore, there are non-profit digital news start-ups have emerged alongside market-driven online natives (e.g., in the US). Wagemans et al. (2016) found that French journalism start-ups are often inspired by US businesses that strive for fact-based, objective journalism. Similarly, for Nordic journalists, US journalism often serves as inspiration. There is a long history of the Americanization of European journalism, for example with norms, routines, and textual conventions previously being adapted to European national contexts (Broersma, 2019). However, Örnebring (2009) argues that, in a Nordic context, "Americanization is largely confined to news format, not news content" (p. 11).

Wagemans et al. (2016) note that most start-ups fail despite good intentions. One explanation for this could be that founders tend to primarily focus on developing their editorial model and spend little time to focus on revenues during the first few years (Naldi & Picard, 2012). In the case of journalistic start-ups, perceived success has been tied to the ability to offer quality content (Pekkala & Cook, 2012). To secure outreach, the start-ups also define success in terms of community, public service and survival rather than money. Therefore, Naldi and Picard (2012, p. 90) use a three-dimensional definition of initial success: 1) ability to attain original expectations, 2) capacity to change the business model, and 3) probability of survival (sustainability). Similarly, Bruno and Kleis Nielsen (2012) argue that a measure of success for digital news start-ups is mere survival. However, the ability to survive can be explained by the ability to deliver a distinct, quality product, operate with a lean organization, have diverse revenues, and target niche audiences that are poorly served by existing media (p. 6).

2.2. Data Journalism as a Form of Entrepreneurial Journalism

Cheruiyot et al. (2019) argue that data journalism is not a purely "journalistic" phenomenon because it is associated with peripheral actors who in various ways complement or expand the work journalists do (p. 2). For example, Nordic data journalists connect with each other through social media, such as the Facebook group "Datajournalistik," at conferences focused on data and journalism (Appelgren, 2016). In Europe, data journalism has been shaped by institutional factors, such as the digitization of public records. These records are increasingly accessible via Application Programming Interface (API) mandated by the PSI Directive (Appelgren & Nygren, 2014) and the increasing supply of tools for collecting, analyzing, and visualizing data. Stalph and Borges-Rey (2018) assert that data journalism has become an asset within legacy organizations over the past decade. However, in this context, the nearby metaphor "robot journalism" has proved to have negative, almost existential, connotations (Lindén & Dierickx, 2019; Willis, 2020).

Based on previous research on boundary work in science, Carlson (2015) suggests a matrix to situate individual studies of boundary work in journalism. The model is based on studies of the separate lines between journalists and non-journalists. In a typical study of participants that move across Carlson's suggested categories of expansion, expulsion, and protection of autonomy, Baack (2018) found that data journalists and other professionals, such as interloping newcomers (e.g., civic technologists), engage in communities of practice, meaning that individuals rarely revolve around sustained and institutionalized engagement (p. 676).

While previous studies on entrepreneurial journalism have focused on pioneering forms of journalism in newsrooms, they have somewhat neglected individual pioneering journalists (Hepp & Loosen, 2019, p. 2). Similarly, even though data journalism is carried out in teams, there is little research on individual data journalists as pioneers. Pioneering journalists have been found to perceive themselves as forerunners who can act as intermediaries and bring together various competencies (Hepp & Loosen, 2019), and the ideal entrepreneurial journalist as embodied in the entrepreneurial discourse is an individual who creates new media products, disrupts traditional media organizations, and generates capital, new products, and even jobs for other journalists (Cohen, 2012, p. 524). In this context, there is an imperative to strive towards upgrading one's own skills, in particular concerning technology and self-promotion. Summarizing the scholarly research on data journalism, Loosen et al. (2017, p. 3) find that scholars are debating if and to what extent data journalism is actually a new reporting style, but the skills of the "technical journalist," such as programming and data journalism, are currently upgrading journalism (Bakker, 2014).

3. Background and Motivating Questions

In this section, we will briefly introduce the Nordic media market and provide the background on the two start-ups. We will then present the motivating questions for the case study.

The media ecosystem in Sweden, Norway, Finland and Denmark has been labelled the Nordic welfare state media model (Syvertsen, Mjøs, Enli, & Moe, 2014). It is characterized by strong public service and decentralized press, with a lot of local and regional newspapers (Hallin & Mancini, 2004). Social media platforms, and

Facebook in particular, have recently become the most frequent source of news for people in this region (Olsen, Solvoll, & Gran, 2018), but there are signs of a general trend away from dependence on platforms (Lindén, 2020). Legacy media is struggling with its business model, but the larger commercial media companies, such as Bonnier, Schibsted, Sanoma and Amedia, are successfully transitioning from mixed revenues—advertisement and subscriptions—to paid digital content (Villi et al., 2019).

The first start-up, Journalism ++ Stockholm, was founded by Swedish-Finnish journalist Jens Finnäs and Swedish journalist Peter Grensund in 2013 (Andén, 2013) in Stockholm. It was part of the European Journalism++ network with chapters in Paris, Berlin, Stockholm, Porto, Amsterdam, and Cologne. Later, the Nordic agency added three new employees. The larger European Journalism++ network was founded in 2011 by Nicolas Kayser-Bril, Pierre Romera, and Anne-Lise Bouyer (Kayser-Bril, 2017). The most famous work from this network was the award-winning project "The Migrants' Files". The European network Journalism++ coordinated this groundbreaking project, which aimed to measure the number of people who died while trying to reach Europe, between 2013 and 2016 (The Migrants' Files, 2014). The network shrank to only two chapters in 2017: Stockholm and Porto.

The second start-up, Journalism Robotics, was first described in 2016 as the product Newsworthy, and later presented as a separate company, Journalism Robotics. Newsworthy is described as a "machine that will find news in data" (Journalism Robotics/Newsworthy, 2020).

J++ Stockholm has been active in the global open-source movement, sharing software code and practices on platforms such as GitHub. Furthermore, it attended key data journalism conferences and seminars, not just in Sweden and Nordic countries but also internationally, such as the European gathering Dataharvest and the US NICAR and Computation + Journalism Symposium conferences. With the emergence of artificial intelligence in journalism, the news media is becoming increasingly dependent on technical expertise (Beckett, 2019), and helping newsrooms understand and make use of technology to produce various forms of data-driven journalism is at the center of the J++ Stockholm and Journalism Robotics/Newsworthy business model. In order to explore how the start-ups negotiated their area of expertise in data journalism over time as they develop the company and their skills, we use two motivational questions: 1) "What are the aims and target groups, i.e., customers, for the two news start-ups?"; and 2) "How do the founders of the news start-ups define their success and failure in navigating the boundaries between journalism and technology?"

4. Methods and Material

This case study is based on a qualitative content analysis of blog posts from the J++ Stockholm and Journalism Robotics/Newsworthy websites and a series of semi-structured interviews with the main founder, Jens Finnäs, and a cofounder, Måns Magnusson, of two news start-ups, Journalism ++ Stockholm and Journalism Robotics/Newsworthy.

Case study research is an empirical process that investigates a contemporary phenomenon, and it is iterative in nature (Yin, 2014). Bryman (2008) argues that it is not always possible to discern the case study type until after a detailed investigation. However, because the case selected for this study is an example of the more general case of a peripheral actor in journalism in the Nordics, we believe that it can be viewed as what Yin (2014) referred to as a representative case.

The interviews with the two founders were conducted in Stockholm and Helsinki, respectively, on November 24, 2016, August 9, 2017, and March 15, November 8, and November 11, 2019. The interviews lasted around 40–50 min each. The interviews were transcribed and analyzed using a thematic and inductive qualitative text analysis approach, where careful reading of transcripts has formed the motivating questions that structure the article. This method has been applied in a flexible way, taking into account the nuances and diversity of the responses. The researchers are positioned as active in the research process; themes do not just emerge (Braun & Clarke, 2006). However, the text analysis is grounded in the subjective meaning of human action, thus preserving the interviewees' subjective point of view (Fereday & Muir-Cochrane, 2006).

To ensure validity, we showed the informants the results of the interview study and corrected any misunderstandings. The corrections were mainly about the order of events in time, the narrative of leaving journalism behind, and the labelling of the companies and products as start-ups. Because the material mainly relates to specific circumstances regarding a small number of individuals, we applied an idiographic approach to the analysis (Bryman, 2008).

For the content analysis, we collected the blog posts available on both companies' web pages. In total, Journalism ++ Stockholm and Journalism Robotics/Newsworthy created 23 blog posts between June 2, 2015 and October 16, 2019 (18 and 5, respectively). The majority of Journalism ++ Stockholm's posts were written by two authors: nine by Jens Finnäs, seven by Leo Wallentin, and the remainder by other staff members. The Journalism Robotics/Newsworthy posts do not include bylines. When coding, the researcher looked for the process of narration in the material, i.e., how the information was offered, withheld or delayed (Gillespie, 2006), or how the company positioned and described itself and its development. We decided to use a qualitative content analysis of company blog posts, where one of the researchers first coded for specific themes in the textual content (data journalism, investigative journalism, technology, non-journalistic practices, success, failure, companies and organizations) and then looked for how these

themes were narrated when promoting the company's aims and activities.

We regard the Journalism ++ Stockholm blog posts as "physical traces of how organizations represent and account for themselves" (Coffey, 2014, p. 367), and the purpose has been to gather information about "sense-making practices" (McKee, 2003, p. 52). The material was read several times and coded (Bryman, 2008). After the initial coding, we found that the blog posts illustrated company development by accounting for the company's activities over time. The results from the content analysis follow the motivating questions and are organized in terms of the intention and language for whom the content was created and for what purpose, as well as the cultural resources the stories draw on and what they aim to accomplish with the stories (Riessman, 1993). All quotes are translated from Swedish.

5. Results

In this section, we will present the results from the content analysis of blog posts and the interviews with the founders of the two digital journalism start-ups.

5.1. Results from the Content Analysis of Blog Posts

An overview of the two blogs forming the basis for the content analysis shows that they utilize different styles. The Journalism ++ Stockholm blog has a personal tone and uses colloquial language with a rich flora of data journalism terms that might be new to non-data journalists. Instead, the Journalism Robotics/Newsworthy blog primarily addresses non-journalists and promotes Newsworthy as a product. However, there are similarities between the blogs in terms of rhetoric. Both blogs frequently use questions at the beginning of texts and paragraphs. Statistics and results from analyses are presented with dialogical language, referring to the reader as "you" and the writer together with the reader as "we," thus guiding the reader in a personal style through complex descriptions of data journalism methods. Both blogs contain images and graphics primarily for the purpose of illustrating relationships found in data.

The first motivational question of our study, as presented earlier, relates to the aims and target groups. We found that the blog posts accounts for company aims by describing development related to data journalism competence and skills. Target groups are mentioned in examples of collaborations with media companies or non-journalistic organizations. Several narratives illustrate the founders' experiences of becoming more skilled at data journalism and developing their company ideas. We therefore begin this section by presenting three quotes that are particularly illustrative of how the company positions itself using personal examples and nestled promotional stories. The first quote provides a historical account of how Journalism ++ Stockholm first was established:

> Four years ago, Jens Finnäs and Peter Grensund received an offer they couldn't refuse: an opportunity to join the Journalism ++ network. They founded Journalism ++ Stockholm and became part of an international network of agencies dedicated to data-driven journalism and newsroom innovation. (Finnäs, 2017)

This quote illustrates how the two journalists entered the international network of data journalists as a successful strategic move. The blog post continues by praising the competence in the international Journalism ++ network, but its main purpose is to mention that key members of the international Journalism ++ network are leaving. This is carefully explained so that it does not seem as if this will do any harm to those remaining in the Journalism ++ network. The blog post then explains that a smaller network will lead to a new exciting era with opportunities to tighten collaborations and become even stronger at producing data journalism.

In the second quote, the presented narrative not only describes the journalistic experience, but functions as a suggestion and an argument for why 'robot journalism,' the main business model of Journalism Robotics/Newsworthy, is valuable to news organizations:

> I began my career as a journalist at the local newspaper Nya Åland. It is not uncommon for local news reporters to feel a bit like a robot. Rewriting police telefax statements as news items about speeding was not very intellectually stimulating. Why couldn't a news robot have done this job? (Finnäs, 2015b)

This quote is followed by calculations of costs that are possible to cut in the newsroom, thus clearly targeting journalists in charge of media houses. The quote aims to motivate senior journalists now in management positions to invest in news automation technology to eliminate mundane tasks. However, in several posts, the start-up founders make it clear that they are not interested in discussing job cuts and replacing journalists with software, but rather the possibilities presented by new technology. The presented experience the founders acquired from having a full career, from the mundane and simple tasks at the junior level as in the quote above, to the present, where expertise in programming makes it possible to reduce such tasks, still implicitly indicates that the mundane experience has been valuable.

The third quote focuses on experience, and is found in the beginning of a Journalism Robotics/Newsworthy blog post on how to create graphs:

> I have been the leader of courses and workshops on how to visualize and tell stories with data intended for journalists and Public Relations professionals for more than ten years. What should you bear in mind when you create a graph? What is the secret behind good data visualization? Here, I have tried

> to reduce everything I have learned in three steps. (Newsworthy, 2019a)

The post appears to be about presenting useful information in graphs, but it implicitly promotes both the Newsworthy product, which provides analytics as data, and the competence of the product's creators. The promotional examples given in the post are primarily non-journalistic and use personal pronouns: "Sales have gone down since *we* changed *our* CEO, or housing prices in *our* municipality are up" (Newsworthy, 2019a). Furthermore, the signal to the reader is that they might not be competent enough in creating graphs. This is visible in other blog posts as well, for example by referring to the audience as having no time for developing skills or journalists notoriously being bad at math. While the quote above is rather humble, the style quickly changes in the blog post. The insights gained from ten years of teaching visualization are summarized as what most journalists do wrong, rather than what journalists can do right, and the author's tone can be interpreted as annoyed or even superior:

> This is by far the most important step. Let me clarify: The most I-M-P-O-R-T-A-N-T. (Newsworthy, 2019a)

The second motivational question of our study is related to critical factors for success. The blog posts discuss the company's success by listing successful projects. On rare occasions, successes are described explicitly, but not by the authors themselves. For example, the following is a press release that indicates success through the quote of a new employee:

> This is very exciting! Newsworthy and J++ [Stockholm] are among Sweden's most innovative media players. (Newsworthy, 2019b)

The post continues with quotes from the founders about how happy they are that the group will now be able to carry out even more skilled data journalistic tasks. Other posts include comments from previous course attendees praising the teaching and what they have learned by attending Journalism ++ Stockholm courses.

The general finding from the analysis of blog posts, however, is that the focus is not on explicit success but rather on presenting the methodology behind successful projects. There is a strong emphasis on the difficulties each project faced. Thus, success is primarily narrated as being transparent with learning new things, as captured in the following quote:

> One of our internal goals at J++ [Stockholm] is for us to work with aspects that are new to us in all our projects. There has to be something innovative in the genre or something that we ourselves have not yet mastered. (Wallentin, 2015)

The tension between technology and journalism and the applied journalistic thinking to non-journalistic areas was visible in the blog posts. We found that substantial effort was made to put future 'robot journalism' in a positive light, listing strengths such as efficiency, endurance and ability to avoid human error. The 'robots' are said to be better at judging quality, for example when interpreting statistics.

In a blog post aimed at promoting an early version of Journalism Newsworthy by describing how 'robot journalism' works, a timeframe is used to impress journalists. Two hours after a public agency releases data, a news automation tool will have analyzed the published data and created 311 reports. In other words, the impossible is now possible. The blog post continues by describing how this process is effortless, as results "drop in from all over the country." This process is also enjoyable:

> We have not only written 1,244 reports (with only two people at the wheel), but we have also created a machine that can create 1,244 reports while we are having a cup of coffee. (Finnäs, 2018)

In this context, mentioning the coffee break signals that it is possible to get large amounts of work done while taking a break. Again, the impossible is now possible. While the blog posts also state that additional human competence is needed to carry out journalistic tasks, and that humans are shaping technological processes, for example that humans have written the news automation software, 'robots' are nevertheless presented as living beings with feelings:

> Carrying out repetitive tasks are one of robot journalism's primary strengths. For the robot, it does not matter if it will get an assignment to write one police report or a thousand. It is just as happy. (Finnäs, 2015b)

Gradually, blog posts increasingly contain non-journalistic examples. Results, or rather methodology behind projects in collaboration with public agencies and unions, shape the message to appeal more to professional but possibly non-journalistic target audiences. Yet the reader is constantly reminded that the products and expertise can be used for journalistic purposes:

> Newsworthy started as an initiative for helping local journalists find news in data. Now, we are obviously also helping elected members in the Teachers' Union understand their world with data....In our news service Newsworthy, we try to build a machine that can answer these questions *en masse* for different types of data. If you just keep the questions at the back of your head the next time you present numbers to your boss, you will have a much better shot at making your voice heard. (Newsworthy, 2019c)

5.2. Results from the Qualitative Interviews with the Start-Up Founders

This section summarizes the interviews with the two founders and is structured around the two motivational questions. The first motivational question, as presented above, concerns the aims and target groups for the two news start-ups. In a report written by Jens Finnäs, he defines the start-up's aim as follows:

> The ambition is to be the leading data journalism freelance agency in Scandinavia, an actor that newsrooms across the Nordic countries can turn to for assistance in data-driven reporting. (Finnäs, 2015a)

In essence, this is the definition of a service provider, and this aim is also reflected in the company's blog posts through their emphasis on describing competence and data journalism methodology. Finnäs explained the logic of the two different corporate structures as two commercial operations with slightly different focuses. According to Finnäs (personal communication, November 11, 2019), Journalism ++ Stockholm is a consultancy firm that helps media companies with various journalistic projects and also trains reporters in data journalism, while Journalism Robotics/Newsworthy has a broader user group, including people in public relations and communication. According to Finnäs, the infrastructure for local content can easily be turned into press releases, personalized emails, or PowerPoint presentations as well as articles.

While the aim of Journalism ++ Stockholm and the reasons for starting the network were explicit in the analyzed blog posts, the interviews revealed that the idea behind Journalism Robotics/Newsworthy emerged from a friendship with statistician Måns Magnusson. Initially, Finnäs had been a freelance data journalist and explored how open public data sources could be used for journalism. Finnäs was friends with a statistics PhD student, Måns Magnusson, who was focusing on machine learning. They were both inspired by US statistician Nate Silver and his start-up FiveThirtyEight, which uses statistical analysis to tell compelling stories.

Before receiving his PhD degree, Magnusson worked as a statistician for the Swedish National Council for Crime Prevention, a popular data source among Swedish data journalists. Magnusson and Finnäs started collaborating professionally, one focusing on hardcore data mining and the other on searching for interesting data that could be used and visualized in journalistic projects. In 2016, their collaboration resulted in Journalism Robotics, which developed news automation services under the brand name Newsworthy. Magnusson (personal communication, November 8, 2019) described the service as "computational storyfinding, going through large amounts of data with statistical methods, finding a needle in the haystack." Jens Finnäs asserts that the start-up's journalistic logic of working with structured data is appealing to more than just journalists:

> Newsworthy adjusts created texts to local contexts and patterns to create a sophisticated software pipeline that autonomously generates texts with very little human effort in terms of adjusting content. We are building the infrastructure to create local content…providing 300 different local news angles instead of one national angle. (Finnäs, personal communication, November 11, 2019)

Apart from producing news content on commission for larger Nordic news organizations, the company also applied for funding from non-profit organizations and foundations, although projects funded in this manner generally were terminated once the funding was gone. One example of this was a project to develop a data mining system for a large newspaper in Tampere, Finland, that failed to gain acceptance in the newsroom. The failure of technology adoptions in newsrooms is a common thread in research (Wagemans & Witschge, 2019).

When the Google-funded Digital News Initiative (DNI) started in 2015, Finnäs and Magnusson decided to apply for funding. They received funding for an idea of news automation based on data sets from Statistics Sweden, a government agency that produces and publishes official statistics. In interviews, Finnäs describes this moment as a defining and critical success factor for the company, and the idea developed into the start-up news service Newsworthy. Overall, Google has been a crucial supporter of data journalism and media innovation (Fanta, 2018; Lindén, 2020). In 2016, Magnusson and Finnäs were involved in an incubator program managed by the Swedish innovation agency Vinnova. This project was plagued by communication problems and what was perceived as less-than-helpful advice: "Don't get me started on that," Magnusson said afterwards (personal communication, November 8, 2019). The main problem was that the funding from this program could not be used to develop their current business model; rather, it could only be used to develop some kind of tangible product. Finnäs and Magnusson struggled to find a solution that Vinnova was willing to fund, and they were asked to "think outside the box" and "be brave." Ultimately, they received no funding.

Reflecting on the project funded by DNI, Finnäs (personal communication, November 11, 2019) believes that they were doing things quite differently today compared to their original pitch. Newsworthy's main function was supposed to be to provide local newsrooms with unique content, and more specifically localized news snippets. Because of the lack of commercial interest among Swedish newsrooms, though, they decided to broaden its services to other customers. Newsworthy landed new customers outside of journalism, such as the Swedish teachers' union, *Lärarförbundet*.

The second motivating question of this study, as presented above, is related to perceived success and failure.

A key part of the perceived success has been to educate journalists and take part in conferences. According

to Finnäs, "Through all these workshops, we have seen that the data skills and Excel skills are very, very low" (personal communication, November 11, 2019). Finnäs has also held a key role in the Nordic Data Journalism Conference and built the repository for entries to NODA Awards as part of attracting new collaborations: "Conferences have been totally invaluable for finding coworkers and customers. For me personally, everything can be traced back to some conference" (Finnäs, personal communication, November 11, 2019).

Finnäs accounts the failures to a lack of interest for the Journalism Robotics/Newsworthy product. The perceived non-response from newsrooms to their offer was disappointing, especially since Finnäs is confident that the tool could help overcome structural obstacles in editorial routines for an industry in which both money and time are scarce.

We found in the interviews that the founders primarily told stories about designing and building the systems, but they somewhat neglected their strategies for addressing target groups to be successful. Finnäs is self-critical when he reflects on how to get people already struggling with information overflow to listen and accept that even more new information could be useful:

> The challenge is not technical, more like organizational. Looking at municipalities, we need to know who will gain and in what way? What are their roles, are they bureaucrats or elected? What topics are interesting and relevant and where do they get their information today? (Finnäs, personal communication, November 11, 2019)

Magnusson is more optimistic, asserting that the low level of interest is only temporary and Newsworthy will be able to focus on journalistic services in the future:

> Since the amount of available data is quickly growing, interesting stories will drown in an unmanageable mass of information. With the help of machine-learning tools such as Newsworthy, journalists can build systems that go through massive amounts of data to find both anomalies and large societal trends. (Magnusson, personal communication, November 8, 2019)

6. Conclusions

Journalism ++ Stockholm and Journalism Robotics/Newsworthy are well-known in the Nordic media market and viewed as pioneers driving change in Nordic news innovation. As such, they are examples of non-traditional actors that matter in journalism (Hermida & Young, 2019). Yet, in this case study, we have found that the company founders are now strategically moving towards non-journalistic customers. With this study, we aim to expand the digital native news start-up research by exploring two digital natives without a direct audience that are working at the periphery of the legacy media, successfully managing to apply journalistic thinking outside of journalistic contexts.

Two motivational questions guided us through our exploration of the digital journalism start-ups' journey. The first motivating question concerned aims and target groups for the two start-ups.

Journalism ++ Stockholm and Journalism Robotics/Newsworthy initially were examples of what Cook and Sirkkunen (2015) denoted as service-oriented digital native sites; they focused on elements of data journalism and had a strategy to sell products to legacy media. In line with García-Avilés et al. (2019) and Bruno and Kleis Nielsen (2012), we also found that they offer alternative products, such as organizing events, and they diversify their revenues with consultancy work and content syndication. To face economic realities, Journalism ++ Stockholm and Journalism Robotics/Newsworthy have gradually become more flexible in terms of their focus on data journalism as a service for legacy media and successfully attracted a few non-journalistic customers.

The founders have not put much effort into researching the needs of either their journalistic or non-journalistic target groups. However, interviews and company blog posts reveal that the founders are very knowledgeable about what can be done with data in a journalistic setting and how to create quality content, an important presumption for news start-up success (Pekkala & Cook, 2012). There is also evidence that they have mastered how to collect and structure data, for example by offering transparent explanations of methodologies, but we did not find explicit offers for attracting customers to the Newsworthy product. The absence of a customer needs analysis beyond the news production process might be an example of what Naldi and Picard (2012) denote as "formal myopia," thus restricting innovation and development beyond the competences and experiences that the founders brought with them into the companies.

The second motivational question revolves around how the companies define their success and failure in navigating the boundaries between journalism and technology.

We argue that the decision to continue with the operations, despite the failure of the larger European J++ network, shows that mere survival, as suggested by Bruno and Kleis Nielsen (2012), was a factor of success. However, training journalists with computational skills has probably been the most important success factor for J++ Stockholm, establishing their competence niche in data journalism. Furthermore, throughout the company's existence, it has successfully produced data journalism for several newsrooms across Scandinavia on commission. This is in line with how Wagemans et al. (2016) argues that core values of journalism and maintaining the professional identity can be part of the main selling point for new forms of entrepreneurial journalism. In general, however, we found that Journalism ++

Stockholm expresses success in a quite negative manner. However, explaining gained insights in blog posts as overcoming failures, a transparent way of expressing skill development may be well-suited to attract data journalistic projects on commission. Traditional media companies were willing to pay the founders for their freelance data journalism content and data journalism training courses in a traditional buyer–seller relationship.

However, we found that the Nordic media companies were reluctant to invest in journalistic products, such as Newsworthy. We argue that the strong emphasis on the contested metaphor 'robot journalism' in the promotional discourse could have been an obstacle. Robot journalism is associated with job loss and thus brings negative connotations to newsroom staff (Lindén & Dierickx, 2019), furthermore, robots are portrayed as more than just automation of mundane tasks, and may thus appear threatening. Therefore, even though journalism forms the basis of the two start-ups, the Journalism Robotics/Newsworthy product, with its emphasis on robotics, may have appeared as challenging, presenting contradictory definitions of the profession (Eldridge, 2018). When trying to sell the product as technology-enhancing journalism, the founders move from the buyer-seller relationship between journalists, as seen when selling data journalism projects on commission, to outsiders, i.e., interloping actors (Belair-Gagnon & Holton, 2018). With the narratives of efficiency and automation, a competing narrative is provided. Reluctance could thus be an example of how legacy media protects autonomy by fending off non-journalists seeking control or who aim to shape journalism (Carlson, 2015).

To conclude, our case study shows that the two start-ups engage heavily in developing technology skills and use the blogs for self-promotion in a manner typical of entrepreneurial journalists (Cohen, 2012). They found a successful narrative to attract data journalistic work on commission acting as journalists, yet become outsiders when selling their product Newsworthy to legacy media. They master the advantage of what Carlson (2016) describes as the metajournalistic discourse regarding cooperation between journalists and technologists by transparently describing failures and methodological challenges in data journalism in their promotional blog posts. However, they have yet to learn how to identify customer needs and communicate what the product can offer in a manner that is appealing to both journalistic and non-journalistic customers. Nevertheless, they attract non-journalistic customers.

Future research could investigate if this type of promotional communication needs to be formulated differently in order to sell the product in non-journalistic contexts and at the same time appeal to the professional logic of journalism. Perhaps the core issue in selling the product Newsworthy is not the fault of the two start-ups, but rather highlights the shortcomings of legacy media to understand why and how they also need to cooperate with peripheral actors.

Acknowledgments

The authors wish to thank Megan Owens for proof reading and the anonymous reviewers for valuable comments. This article is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media). The results of this article reflect only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

Conflict of Interests

The authors declare no conflict of interests.

References

Andén, A. (2013, January 23). Vad är affärsnyttan med datajournalistik, Jens Finnäs? [What is the business value of data journalism, Jens Finnäs?]. *Medievärlden*. Retrieved from https://www.dagensmedia.se/nyheter/medievarldens-arkiv/vad-ar-affarsnyttan-med-datajournalistik-jens-finnas

Appelgren, E. (2016). Data journalists using Facebook. *Nordicom Review*, *37*(1), 156–169.

Appelgren, E., & Nygren, G. (2014). Data journalism in Sweden: Introducing new methods and genres of journalism into "old" organizations. *Digital Journalism*, *2*(3), 394–405.

Baack, S. (2018). Practically engaged: The entanglements between data journalism and civic tech. *Digital Journalism*, *6*(6), 673–692.

Bakker, P. (2014). Mr. Gates returns: Curation, community management and other new roles for journalists. *Journalism Studies*, *15*(5), 596–606.

Beckett, C. (2019). *New powers, new responsibilities: A global survey of journalism and artificial intelligence*. London: The London School of Economics and Political Science.

Belair-Gagnon, V., & Holton, A. E. (2018). Strangers to the game? Interlopers, intralopers, and shifting news production. *Media and Communication*, *6*(4), 70–78.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Broersma, M. (2019). Americanization or the rhetoric of modernity: How European journalism adapted US norms, practices and conventions. In K. Arnold, P. Preston, & S. Kinnebrock (Eds.), *The Handbook of European communication history* (pp. 403–419). New York, NY: Wiley.

Bruno, N., & Nielsen, R. (2012). *Survival is success: Journalistic online start-ups in western Europe*. Oxford: Reuters Institute for the Study of Journalism.

Bryman, A. (2008). *Social science research methods* (3rd ed.). Oxford: Oxford University Press.

COGITATIO

Carlson, M. (2015). Introduction: The many boundaries of journalism. In S. C. Lewis & M. Carlson (Eds.), *Boundaries of journalism* (pp. 1–18). New York, NY: Routledge.

Carlson, M. (2016). Metajournalistic discourse and the meanings of journalism: Definitional control, boundary work, and legitimation. *Communication Theory*, *26*(4), 349–368.

Cheruiyot, D., Baack, S., & Ferrer-Conill, R. (2019). Data journalism beyond legacy media: The case of African and European civic technology organizations. *Digital Journalism*, *7*(9), 1–15.

Coddington, M. (2015). The wall becomes a curtain: Revisiting journalism's news–business boundary. In S. C. Lewis & M. Carlson (Eds.), *Boundaries of journalism* (pp. 67–82). New York, NY: Routledge.

Coffey, A. (2014). Analysing documents. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 367–380). London: SAGE.

Cohen, N. S. (2012). Entrepreneurial journalism and the precarious state of media work. *South Atlantic Quarterly*, *114*(3), 513–533.

Cook, C., & Sirkkunen, E. (2015). What's in a niche? Exploring the business model of online journalism. *Journal of Media Business Studies*, *10*(4), 63–82.

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Cambridge, MA: Harvard University Press.

Eldridge, S. A., II. (2018). *Online journalism from the periphery: Interloper media and the journalistic field*. New York, NY: Routledge.

Fanta, A. (2018). The publisher's patron: How Google's News Initiative is re-defining journalism. *European Journalism Observatory*. Retrieved from https://en.ejo.ch/digital-news/the-publishers-patron

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, *5*(1), 80–92.

Finnäs, J. (2015a). *Report from NICAR 2015*. Helsinki: Finnmedia. Retrieved from http://www.medialiitto.fi/files/3132/Jens_Finnas_matka-apuraha_NICAR-konferenssiin_2015.pdf

Finnäs, J. (2015b, June 29). Därför behöver vi fler robotjournalister [That is why we need more robot jouranlists]. Retreived from jplusplus.org/sv/blog/darfor-behover-vi-fler-robotjournalister

Finnäs, J. (2017, April 11). Journalism++ enters a new era. Retrieved from jplusplus.org/sv/blog/journalism-enters-a-new-era

Finnäs, J. (2018, January 19). 1 244 lokala nyheter på 2 personer och 24 timmar [1244 local news reports created by 2 persons in 24 hours]. Retrieved from jplusplus.org/sv/blog/1244-lokala-nyheter-pa-2-personer-och-24-timmar

García-Avilés, J. A., Carvajal-Prieto, M., Arias, F., & De Lara-González, A. (2019). Journalists' views on innovating in the newsroom: Proposing a model of the diffusion of innovations in media outlets. *The Journal of Media Innovations*, *5*(1), 1–16.

Gillespie, M. (2006). Analysing media texts. In M. Gillespie & J. Toynbee (Eds.), *Narrative analysis* (pp. 79–117). Maidenhead: Open University Press.

Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge: Cambridge University Press.

Harlow, S., & Salaverría, R. (2016). Regenerating journalism: Exploring the "alternativeness" and "digitalness" of online-native media in Latin America. *Digital Journalism*, *4*(8), 1001–1019.

Hepp, A., & Loosen, W. (2019). Pioneer journalism: Conceptualizing the role of pioneer journalists and pioneer communities in the organizational re-figuration of journalism. *Journalism*. https://doi.org/10.1177/1464884919829277

Hermida, A., & Young, M. L. (2019). From peripheral to integral? A digital-born journalism not for profit in a time of crises. *Media and Communication*, *7*(4), 92–102.

Journalism Robotics/Newsworthy. (2020). Statistik innehåller nyheter: Vi hittar dem åt dig. [Statistics contain news: We find them for you]. *Newsworthy*. Retrieved from https://www.newsworthy.se/sv

Kayser-Bril, N. (2017). Things learned in 5 years of journalism++. *Nkb*. Retrieved from https://blog.nkb.fr/things-learned-jpp

Knight, M. (2015). Data journalism in the UK: A preliminary analysis of form and content. *Journal of Media practice*, *16*(1), 55–72.

Konieczna, M., & Powers, E. (2017). What can nonprofit journalists actually do for democracy? *Journalism Studies*, *18*(12), 1542–1558.

Lindén, T. C. G., & Dierickx, L. (2019). Robot journalism: The damage done by a metaphor. *Unmediated: Journal of Politics and Communication*, *2*, 152–155.

Lindén, T. C. G. (2020). *Silicon Valley och makten över medierna* [Silicon Valley and the power over media]. Gothenburg: Nordicom.

Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2017). Data-driven reporting: An on-going (r) evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*. https://doi.org/10.1177/1464884917735691

McKee, A. (2003). *Textual analysis: A beginner's guide*. London: SAGE.

Naldi, L., & Picard, R. G. (2012). "Let's start an online news site": Opportunities, resources, strategy, and formational myopia in startups. *Journal of Media Business Studies*, *9*(4), 69–97.

Nee, R. C. (2013). Creative destruction: An exploratory study of how digitally native news nonprofits are innovating online journalism practices. *International Journal on Media Management*, *15*(1), 3–22.

Newsworthy. (2019a, June 4). Hur gör man ett bra diagram? [How to create a graph?]. Retrieved

from https://medium.com/@contact_76790/hur-g%C3%B6r-man-ett-bra-diagram-e1cc69da9d9a
Newsworthy. (2019b, June 4). Newsworty rekryterar från BBC [Newsworthy recruits from the BBC]. Retrieved from https://medium.com/@contact_76790/newsworthy-rekryterar-från-bbc-c2399e7b23fc
Newsworthy. (2019c, October 16). En ranking [A ranking]. Retrieved from https://medium.com/@contact_76790/en-ranking-290-powerpoint-presentationer-383becdb38d1
Olsen, R., Solvoll, M., & Gran, A.-B. (2018). *Digitalisering av lokal mediebruk* [Digitization of local media use]. Oslo: BI Centre for Creative Industries.
Örnebring, H. (2009). *Comparative European journalism: The state of current research*. Oxford: University of Oxford.
Pekkala, P., & Cook, C. E. (2012). Sustaining journalistic entrepreneurship. In E. Sirkkunen & C. Cook (Eds.), *Chasing sustainability on the net: International research on 69 journalistic pure players and their business models* (pp. 110–135). Tampere: Tampere Research Centre for Journalism, Media and Communication.
Powers, M., & Zambrano, S. V. (2016). Explaining the formation of online news start-ups in France and the United States: A field analysis. *Journal of Communication*, *66*(5), 857–877.
Raviola, E. (2019). Just like any other business or a special case? Framing excess in a Swedish newspaper group. In B. Czarniawska & O. Löfgren (Eds.), *Overwhelmed by overflows?* (pp. 96–110). Lund: Lund University Press.
Riessman, C. K. (1993). *Narrative analysis* (Vol. 30). Newbury Park, CA: SAGE.
Stalph, F., & Borges-Rey, E. (2018). Data journalism sustainability: An outlook on the future of data-driven reporting. *Digital Journalism*, *6*(8), 1078–1089.
Syvertsen, T., Mjøs, O. J., Enli, G. S., & Moe, H. (2014). *The media welfare state: Nordic media in the digital era*. Ann Arbor, MI: University of Michigan Press.
Tenor, C. (2019). Logic of an effectuating hyperlocal. *Nordicom Review*, *40*(s2), 129–145.
The Migrants' Files. (2014). *The migrants' files*. Retrieved from http://www.themigrantsfiles.com
Usher, N. (2017). Venture-backed news start-ups and the field of journalism: Challenges, changes, and consistencies. *Digital Journalism*, *5*(9), 1116–1133.
Villi, M., Grönlund, M., Linden, C. G., Lehtisaari, K., Mierzejewska, B., Picard, R. G., & Rõpnack, A. (2019). "They're a little bit squeezed in the middle": Strategic challenges for innovation in US metropolitan newspaper organisations. *Journal of Media Business Studies*. https://doi.org/10.1080/16522354.2019.1630099
Wagemans, A., Witschge, T., & Deuze, M. (2016). Ideology as resource in entrepreneurial journalism: The French online news startup Mediapart. *Journalism Practice*, *10*(2), 160–177.
Willis, H. (2020). Journalism on autopilot: The upside and downside of computer-generated stories. *Quill Magazine*. Retrieved from https://www.quillmag.com/2020/01/14/journalism-on-autopilot-the-upside-and-downside-of-computer-generated-stories
Wagemans, A., & Witschge, T. (2019). Examining innovation as process: Action research in journalism studies. *Convergence*, *25*(2), 209–224.
Wallentin, L. (2015, June 15). En faktagranskningsmetod för datajournalistik [A factchecking method for data journalism]. Retreived from jplusplus.org/sv/blog/en-faktagranskningsmetod-for-datajournalistik
Yin, R. (2014). *Case study research design and methods* (5th ed.). Thousand Oaks, CA: SAGE.

About the Authors

Ester Appelgren is an Associate Professor of Journalism and the Head of School of Natural Sciences, Environmental Studies and Technology at Södertörn University in Sweden. She as a PhD in Media Technology from the Royal Institute of Technology, Sweden. Her research interests focus on data journalism and digital integrity. During 2011–2014, Ester managed an action research-oriented project on data journalism, funded by the Swedish Innovation Agency (Vinnova).

Carl-Gustav Lindén is Associate Professor at the University of Helsinki and a Researcher in the intersection between journalism, media and technology with a focus on innovation and business models. He has led a large Finnish research project developing Natural-Language Generation systems for media (www.immersiveautomation.com) and is currently involved in creating new Artificial Intelligence tools for newsrooms working in lesser resourced languages (www.EMBEDDIA.eu).