How to Apply Privacy by Design in OSINT and big Data Analytics?

Jyri Rajamäki[1, 2] and Jussi Simola[2]
[1]Laurea University of Applied Sciences, Espoo, Finland
[2]University of Jyväskylä, Finland
jyri.rajamaki@laurea.fi
jussi.hm.simola@student.jyu.fi.

Abstract: In a world where technology grows exponentially, more information is available to us every day. States and their governments have collected information on their citizens for a long time now. On the other hand, people give out more and more personal information voluntarily through social media. Information available on the Internet is easier to analyze with modern technologies and the original source of information is also easier to track down. Information is available to all of us and that information can be used to investigate personal data, defeat competitors in a corporate world, solve crimes or even win wars. This study analyses open source intelligence (OSINT) and big data analytics (BDA) with the emphasis on cyber reconnaissance and how personal security is part of that entity. The main question is how privacy manifests itself as part of OSINT and BDA. At the same time the study analyses how law enforcement authorities can act so that their reconnaissance actions would be publicly approved. The study uses case study methodology by gathering a comprehensive list of sources for the theory section. The theoretical framework consists of Privacy by Design approach and privacy questions with regard to surveillance, and the General Data Protection Regulation (GDPR) and the Directive 2016/680 'on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data' act as a legal framework. The empirical case dealing with maritime surveillance, explores OSINT and BDA privacy challenges in the MARISA project. The overall target of the paper is to accelerate the discussion on the serious problem of privacy breach that may lead to restrictions of individual liberty and erosion of our society's foundations of trust.

Keywords: privacy by design, OSINT, big data analytics, maritime surveillance

1. Introduction

New surveillance technologies became omnipresent in our everyday live although surveillance has a bad reputation in most countries (Krempel & Beyerer, 2014). Open source intelligence (OSINT) is intelligence collected from publicly available sources, including the Internet, newspapers, radio, television, government reports and professional and academic literature (Glassman & Kang, 2012), and OSINT is being extensively used by local and national law enforcement authorities (LEAs), intelligence agencies and the military. An important aspect of LEAs use of OSINT is social media, which aggregate huge amounts of data generated by users which are in many cases identified or identifiable (Staniforth, 2016): "When combined with other online and stand-alone datasets, this contributes to create a peculiar technological landscape in which the predictive ability that is Big Data Analytics (BDA) has relevant impact for the implementation of social surveillance systems." BDA of OSINT requires the rigorous review and potential overhaul of existing intelligence models and associated processes, however, LEAs must always ensure that their access and use of publicly available information is within national and international legal frameworks (Staniforth, 2016).

This case study, carried out by Yin's (2009) framework, researches privacy issues of the MARitime Integrated Surveillance Awareness (MARISA) project in maritime surveillance domain. Maritime surveillance is essential for creating maritime awareness, in other words 'knowing what is happening at sea'. Integrated maritime surveillance is about providing authorities interested or active in maritime surveillance with ways to exchange information and data. Support is provided by responding to the needs of a wide range of maritime policies-irregular migration/border control, maritime security, fisheries control, anti-piracy, oil pollution, smuggling etc. Also the global dimension of these policies is addressed, e.g. to help detect unlawful activities in international waters. Sharing data will make surveillance cheaper and more effective. Currently, EU and national authorities responsible for different aspects of surveillance collect data separately and often do not share them. As a result, the same data may be collected more than once. The European Commission and EU/EEA members develop a common information-sharing environment (CISE) that integrates existing surveillance systems and networks and gives all relevant authorities access to the information they need for their missions at sea. CISE will make different systems interoperable so that data can be exchanged easily through the use of modern technologies.

The General Data Protection Regulation (GDPR) and/or the Directive 2016/680 'on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data' regulate processing of personal data in maritime surveillance. Privacy by Design (PbD) is one of the key requirements in both regulations. The purpose of this paper is to understand and build explanation what privacy means with regard to OSINT and BDA. The research questions is: How we can understand PbD in regard to the MARISA services? From the MARISA project's point of view, the target of this case study is to provide research findings for the MARISA project's second phase, in which the already developed MARISA services will be revised and enhanced, and additional services will be included. The overall target of the paper is to accelerate the discussion on the serious problem of privacy breach that may lead to restrictions of individual liberty and erosion of our society's foundations of trust.

2. Theoretical framework

2.1 Privacy by design

Privacy by design (PbD) is an approach to systems engineering approach intended to ensure privacy protection from the earliest stages of a project and to be taken into account throughout the whole engineering process, not just in hindsight. The PbD concept is closely related to the concept of privacy enhancing technologies (PET) published in 1995 (Hustinx, 2010). PbD framework was published in 2009 (Cavoukian, 2011).The concept is an example of value sensitive design that takes human values into account in a well-defined manner throughout the whole process. PbD is one of the key requirements in the European Data Protection Reform beings included in GDPR and Directive 2016/680. The GDPR also requires Privacy by Default, meaning that the strictest privacy settings should be the default.

According to Antignac and Le Métayer (2014), PbD related research has focused on technologies and components rather than methodologies and architectures. They advocate that PbD should be addressed at the architectural level and be associated with suitable methodologies, among other benefits, architectural descriptions enable a more systematic exploration of the design space. In addition, because privacy is intrinsically a complex notion that can be in tension with other requirements, they believe that formal methods should play a key role in this area (Antignac & Le Métayer, 2014). Kung (2014) continues the importance of architecture in designing a PbD system and provides an overview on how architectures are designed, analysed and evaluated, through quality attributes, tactics and architecture patterns. Kung also specifies a straw man architecture design methodology for privacy and present PEAR (Privacy Enhancing ARchitecture) methodology. Martin and Kung (2018) posit that for PbD to be viable, engineers must be effectively involved and endowed with methodological and technological tools closer to their mindset, and which integrate within software and systems engineering methods and tools.

In many respects, original PbD framework has been criticized as being a vague concept. To make its underlying goals more concrete, Colesky Hoepman and Hillen (2016) propose a more specific privacy design strategy: 1) minimize: only collect that data which is strictly necessary, and remove that which no longer is; 2) hide: encrypt, pseudonymize, and take other measures that protect and obscure links between elements of data and their source; 3) abstract: reduce the granularity of data collected; combine or aggregate data from multiple sources so that the sources are no longer uniquely identifiable; 4) separate: store and access data only where it is used; process data at the source instead of centrally; 5) inform: explain to data subjects how their personal data is processed, and how profiles and automated decision-making based on their personal data work. A subject can only provide valid consent to data processing if they understand how their data is being processed; 6) control: allow data subjects to provide and revoke consent to process, and to access, correct, and delete their provided and derived data: 7) enforce: build technical and organizational measures that ensure the design decisions taken with regard to privacy are actually implemented, and log the actions of the systems; and 8) demonstrate: document, audit, and report on the operational and PbD processes. The first four strategies are more focused on data and the last four are about policies and the surrounding processes. Given these strategies, the PbD process could then ideally be implemented as follows (Van Aubel, et al., 2018): "look at each project requirement, figure out what potential privacy impacts it has, and apply strategies to mitigate those impacts". This iterative process should be repeated as the design becomes more detailed (Van Aubel, et al., 2018).

2.2 Privacy and surveillance

The PARIS (PrivAcy pReserving Infrastructure for Surveillance) project (2013-2015) defined and demonstrated a methodological approach for the development of a surveillance infrastructure which enforces the right of citizens for privacy, justice and freedom. The project took into account the evolving nature of such rights, since aspects that are acceptable today might not be acceptable in the future. It also included the social and ethical nature of such rights, since the perception of such rights varies over time and in different countries. Its methodological approach was based on two pillars: 1) a theoretical framework for balancing surveillance and privacy/data protection which fully integrates the concept of accountability; and 2) an associated process for the design of surveillance systems which takes from the start privacy (i.e. Privacy-by-Design) and accountability (i.e. Accountability-by-Design).

Koops (2013) concerns procedural issues of OSINT in police investigations and investigates criminal-procedure law in relation to open source data gathering by the police. He studies the international legal context for gathering data from openly accessible and semi-open sources, including the issue of cross-border gathering of data. This analysis is used to determine if investigating open sources by the police in the Netherlands is allowed on the basis of the general task description of the police, or whether a specific legal basis and appropriate authorization is required for such systematic observation or intelligence. The line between espionage and OSINT can be very thin, therefore caution and double-checking are advised before conducting OSINT activities (Hribar, et al., 2014).

Koops, Hoepman and Leenes (2013) considers the challenge of embedding PbD in OSINT carried out by law enforcement. Ideally, the technical development process of OSINT tools is combined with legal and ethical safeguards in such a way that the resulting products have a legally compliant design, are acceptable within society, and at the same time meet in a sufficiently flexible way the varying requirements of different end-user groups. They use the analytic PbD framework and they discusses two promising approaches, revocable privacy and policy enforcement language. The approaches are tested against three requirements that seem suitable for a 'compliance by design' approach in OSINT: purpose specification; collection and use limitation and data minimization; and data quality (up-to-datedness). For each requirement, they analyze whether and to what extent the approach could work to build in the requirement in the system. They demonstrates that even though not all legal requirements can be embedded fully in OSINT systems, it is possible to embed functionalities that facilitate compliance in allowing end-users to determine to what extent they adopt PbD approach when procuring an OSINT platform, extending it with plug-ins, and fine-tuning it to their needs. Therefore, developers of OSINT platforms and networks have a responsibility to make sure that end-users are enabled to use PbD, by allowing functionalities such as revocable privacy and a policy enforcement language (Koops, et al., 2013). Even though actual end-users have a responsibility of their own for ethical and legal compliance, it is important to recognize that it is questionable whether all responsibility for a proper functioning and use of OSINT platforms can be ascribed to the end-users; and some responsibility for a proper functioning of OSINT framework in practice also lies with the developers of the platform and individual components (Guest Editorial, 2013).

3. Open source related MARISA services

The MARISA toolkit provides a suite of services to correlate and fuse various heterogeneous and homogeneous data and information from different sources, including Internet and social networks. MARISA also aims to build on the huge opportunity that comes from using the open access to big data for maritime surveillance: the availability of large to very large amounts of data, acquired from various sources ranging from sensors, satellites, open source, internal sources and of extracting from these amounts through advanced correlation improves knowledge. The first phase MARISA service description document (MARISA, 2018) defines three open source related services: Twitter service, OSINT service and GDELT service. Next we will present those services.

3.1 Twitter service

Twitter is a popular and widely used social media platform for microblogging, or broadcasting short messages. Twitter has hundreds of millions of users worldwide, and they broadcast over every day 500 million messages, known as tweets that may include text, images, and links (Glasgow, 2015). In crisis management, Twitter can act as a human sensor network for real-time event detection, but little attention has been paid to applying text mining and natural language processing techniques to monitor events in a multilingual setting and most of the work focusses on one single language only (Zielinski, 2013).

MARISA Twitter service enables access to open source social media information. Twitter is selected because its users are fast at creating content and an application program interface (API) is available. Many publications in the field of natural language processing are done using Twitter. MARISA Twitter service will read tweets via the Twitter search API. The input is the Area of Interest (AOI), which contains a geolocation (as point in lat/lon coordinates and a radius in km or miles) and the period of time. Each tweet is first analyzed for its language and then tokenized. A special classifier with a language and domain dependent model will assess the relevance of the tweet in this context (domain, use case). The result will be an instance of the Risk class defined in CISE containing a list of assessed tweets with their relevance exposed in the attributes *RiskProbability, RiskSeverity* and *RiskLevel*. MARISA Twitter service won't correlate position information extracted from social media with known ship positions, as it will just give away all identified risks in a given area. An example may be illegal immigration. For the trained use case of illegal immigration, if the classifier delivers a relevance of i.e. 0.9, there is a high probability that the tweet is about a real immigration event. So the *RiskProbability* is frequent (01), the *RiskSeverity* is catastrophic (01) related to possible death of immigrants and the derived *RiskLevel* is high (01). (MARISA, 2018)

The first function block *ReadTweets* consists of two parallel threads. The thread containing the functions *BuildRequest* and *SendRequest* handles the interpretation of the function argument (AOI) and the construction of the HTTP request for *TwitterSearch* API. The second thread in *ReadTweets* processes the HTTP response of the *TwitterSearch* API. If there are no tweets in the response an *EmptyResponse* object is built and the flow ends. Otherwise the tweets list is handed over to the *AnalyseTweet* function. In the *AnalyseTweet* function the critical operation is to detect the language of a tweet. All subsequent operations are dependent of the correct identification of the language of the short message. If the language is not successfully identified, an *EmptyResponse* object is built with an appropriate error code as return state. Upon successful language detection the tweet is tokenized with a special tokenizer which respects all the special controls and characteristics of a tweet. The tokenized tweets enter the *ClassifyTweet* function. This function's building blocks *FindClass* and *AssignRelevance* are based on a *DeepLearning* concept using paragraph vectors and their vector space similarity characteristics. So two tweets are similar, if their corresponding (paragraph) vectors enclose a small angle in a multi-dimensional space (around 500 dimensions). After each tweet is processed, a *Response* object is built, containing the classification result, and the flow ends. (MARISA, 2018)

3.2 GDELT service

The Global Database of Events, Language, and Tone (GDELT) is a CAMEO-coded dataset containing geo-located events with global coverage from 1979 to the present. The data are collected from news reports throughout the world and the dataset provides daily coverage on the events found in news reports published on that day. In 2015, datasets Mentions and Global Knowledge Graph (GKG) were added to GDELT. The Mentions table records the network trajectory of the story of each event in flight through the global media system while the GKG table expands GDELTs ability to quantify global human society beyond cataloging physical occurrences towards actually representing all of the latent dimensions, geography, and network structure of the global news. Today, GDELT is a real time database of global human society for open research which monitors the world's broadcast, print, and web news, creating a free open platform for computing on the entire world containing three data tables: Event, Mentions and GKG while most researches are based only on the Event table (Chen, et al., 2016). GDELT archives an exhaustive collection of available online news sources in more than 100 languages (Guo & Vargo, 2017).

MARISA GDELT service integrates open-source data from GDELT project into MARISA. It filters the results using natural language processing in order to identify possible events related to maritime domain, such as naval incidents, piracy events, and pollution events (MARISA, 2018).

Satellite data represent major value adding maritime surveillance information outside the coastal systems coverage. Social media information such as Twitter provides no adjunct value offshore far from ports where online news sources based GDELT data are more relevant, including maritime events such as emergencies like sinking ships, collisions at sea, or information related to the conflicts in defending territorial waters. MARISA proves the capability and potentially the exportability to other areas and topics, further to the ones dealt with in the project. (MARISA, 2018)

OpenGeo Suite Web Feature Service is integrated in the MARISA toolkit that filters relevant events and news from GDELT data. The function exploits big data information extraction techniques from non-conventional sources. Via the MARISA toolkit, the user accesses OSINT data and reports relevant information to the maritime domain, classified per event type and with references, to enrich the maritime picture within the selected AOI. Current GDELT database queries have no correlation with vessels, but news relevant to a certain AOI will be extracted from the database and made available to the MARISA toolkit for a potential contextual analysis. (MARISA, 2018)

3.3 OSINT service

MARISA OSINT service involves the collection, analysis and use of data from open sources for intelligence purposes. It exploits existing open source solutions for social media data stream integration, especially Twitter web crawlers in order to provide capabilities for discovering of alert of any kind of illegal activities in the maritime environment. Multilingual investigations into social media, based mainly on the ability to identify geo-located information, allow to associate the OSINT information with more closely related to the marine environment information (e.g. vessels, sea condition, pollution risks) and then generate an improved maritime picture that meets cross-border requirements of the MARISA project. Depending on the search type that the user wants to perform (based on geographical locations or coordinates, dates or specific keywords that will be configured in a specific phase of the usage of the service) different API services have been applied inside the service code. OSINT service can search, identify and merge relevant multilingual events that can be considered as input to generate alert/incident/tracks. (MARISA, 2018)

MARISA OSINT service receives parameter from the configuration (e.g. AOI, keywords), and on the basis of this, solves possible conflicts, receives tweets and after analysis propagate alarms as following: 1) *TwitterRetriever* provides a sort of orchestration of all service's components; 2) *OrganizerParmas_API* valuates if there are conflicts or inconsistencies between the input parameters that would lead to a negative search result; 3) *REST_APIsInvoker* evaluates which representational state transfer API shall invoke, merge or sum the results; 4) *LanguageDetection* evaluates suitable Twitter APIs to invoke, merge or sum the results; and 5) *AlarmPropagator* informs if there are tweets coming from AOI and provides the link for retrieving. Depending on the search type the user wants to perform (based on geographical locations or date), different API services are applied inside the service code. Due to sophisticated combinations of criteria, Twitter service can search and identify the set of relevant tweets that can be considered as alert/incident and from which the list of coordinates can be extracted. (MARISA, 2018)

4. Privacy challenges of MARISA OSINT and BDA services

The MARISA toolkit was built on the top of a big data infrastructure that provides the means to collect external data sources and operational systems products and to organize and exploit all the incoming data as well as all the data produced by the various services. Next we look privacy challenges in four different dimensions of BDA: data generation, data analysis, use of data, and infrastructure behind data.

4.1 Data generation

Data generation can be classified into active data generation and passive data generation: active data generation means that the data owner will give the data to a third party, while passive data generation refers to the circumstances that the data are produced by data owner's online actions (e.g., browsing) and the data owner may not know about that the data are being gathered by a third party (Jain, et al., 2016).

The MARISA Toolkit has two relevant data sources: data coming from the sensors, and data coming from open sources. With regard to data coming from the sensors*,* these sensors are embodied in the operational environment of the Legacy Systems. Here Legacy Systems mean the previously existing end-users Maritime Surveillance systems in the National/Regional Coordination Centres or Coastal Stations to which MARISA Toolkit must establish some kind of communications. In these environments, owned by Participating Member State governmental entities, we can suppose that the data are used on the basis of need-to-know and need-to-share. Examples of those data from heterogeneous sources are radar and AIS tracks, AIS data validation, near real-time satellite detections and heat maps, integration of maps of most used routes (density maps) and traffic patterns, search and rescue risk maps, fusion of surveillance pictures information from end-users operational environments.

MARISA services include three services (Twitter service, GDELT service and OSINT service) that collect open source information. Their main target is to extract and integrate maritime related safety and security events. OSINT service mainly collects its information via Twitter service and DGELT service. From data collection point of view, MARISA GDELT service may not have privacy concerns because professional journalists should have taken that issue into account when making news. However other ethical issues may arise, for example wealthier countries not only continue to attract most of the world news attention, they are also more likely to decide how other countries perceive the world (Guo & Vargo, 2017).

ln Twitter, several technical features and tweet-based social behaviors occur that might compromise privacy. Tweets are complex objects that, in addition to the message content, have many pieces of associated metadata, such as the username of the sender, the date and time the tweet was sent, the geographic coordinates the tweet was sent from if available, and much more (Glasgow, 2015). “Most metadata are readily interpretable by automated systems, whereas tweet message content may require text processing methods for any automated interpretation of meaning” (Glasgow, 2015). “Direct Messages” are the private side of Twitter and "retweeting" is directly quoting and rebroadcasting another user's tweet. Someone might unintentionally or intentionally retweet private tweet to a public forum. Other behaviors include mentioning another user in one's tweet that is, talking about that user. According to Rumbold and Wilson (2018), when one puts any information in the public domain—whether intentionally or not—one does not waive one’s right to privacy, but one can only waive one’s right to privacy by actually waiving it.

4.2 Data analytics

Big data may be analyzed by artificial intelligence (AI). Machine learning (ML), a branch of AI, can provide detailed, personalized characteristics of an individual and prediction of his or her future behavior (Moallem, 2019). According to Wójtowicz and Cellary (2019), one of the most important carateristics of BDA is the paradigm shift, in which instead of discovering knowledge by searching for causality, one can discover it by searching for correlation: it is possible via BDA to learn with high propability what is happening, and even what will happen, but not why it happens or why it will happen. If a human programmer writes a program, another human programmer may inspect program code and find possible errors, but if a neural network is trained by peta-bytes of data, nobody is able to check whether a particular prediction is correct or not (Moallem, 2019).

Algorithms tell computers step by step how to solve a certain problem. However, predictive algorithms are often themselves unpredictable (Wójtowicz & Cellary, 2019). According to Rahman (2017), the first problem comes from algorithmic bias—AI algorithms being a reflection of the programmers’ biases—may possibly give rise to the risk of false alerts by AI surveillance systems thus resulting in wrongful profiling and arrest; and the second problem is that AI profiling systems utilise historical data to generate lists of suspects for the purposes of predicting or solving crimes. ML techniques including neural networks run in two phases (the training phase and the prediction phase) and the quality of predictions is absolutely dependent on examples used for the training phase. ML systems are only as good as the data sets that the systems trained and worked with (Rahman, 2017).

4.3 Use of data

Data analysis does not directly touch the individual and may have no external visibility. An important ethical issue comes with automated policing. Automated discrimination is possible when augmented surveillance becomes more common. It intersects with the technical issues of unintended biases in algorithms and big data that could skew analyses generated by AI systems (Rahman, 2017). If a person is wrongly qualified as a potential terrorist, the consequences may be very severe (Wójtowicz & Cellary, 2019). If BDA provides predictions with 99% accuracy, wrong predictions would concern over 5 million people in the EU, which population is 508 million. Big Data used by law enforcement will increase the chances of certain tagged people to suffer from adverse consequences without the ability to get back or even having knowledge that they are being discriminated (Matturdi, et al., 2014).

4.4 Infrastructure behind data

Data analytics requires not just algorithms and data but also physical platforms where the data are stored and analysed. Cloud computing is currently the most economic option of providing computing power and storage capacity, and privacy assurance can be successfully deployed in private clouds. Although stored data are encrypted and advances in homomorphic encryption, there is no prospect of commercial systems being able to maintain this encryption during real-time processing of large datasets (Wójtowicz & Cellary, 2019). The security and privacy for big data is not different from security and privacy research in general (Nelson & Olovsson, 2016).

5. Summary and conclusion

According to the European Data Protection Reform, PbD is a mandatory approach in maritime surveillance context. Although PbD as a concept is becoming well-known, it turns out that there is not much standardization in how to actually apply it, especially by security authorities. This paper explores privacy challenges in the MARISA project and tries to accelerate the discussion on the serious problem of privacy breach that may lead to restrictions of individual liberty and erosion of our society's foundations of trust. Current academic arguments are shifting the focus of privacy concerns from data collection to data analytics and data use, and there are scholars requiring “algorithmic accountability” (Broeders, et al., 2017). It can therefore be expected that the legal requirements concerning OSINT and BDA may develop into this direction.

One very important issue is who watches the watchers (political issue) and how this can be carried out (technical issue). Utilizing BDA in the security domain requires intensive oversight (Broeders, et al., 2017). However, BDA is often a “black box”, and more research is needed, especially in the phase of the analysis: selecting the algorithms, data sources and categorization, assigning weight to various data, etc.

Acknowledgements

Acknowledgement is paid to the MARISA Maritime Integrated Surveillance Awareness project. This project is funded by the European Commission through the Horizon 2020 Framework under the grant agreement number 740698. The sole responsibility for the content of this paper lies with the authors. It does not necessarily reflect the opinion of the European Commission or of the full project. The European Commission is not responsible for any use that may be made of the information contained therein.

References

Antignac, T. & Le Métayer, D., 2014. Privacy by Design: From Technologies to Architectures. In: Privacy Technologies and Policy. Cham: Springer, pp. 1-17.

Broeders, D. et al., 2017. Big data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data. Computer Law & Security Review, Volume 33, pp. 309-323.

Cavoukian, A., 2011. Privacy by Design: The 7 Foundational Principles, Ontario: Information and Privacy Commissioner of Ontario.

Chen, K., Qiao, F. & Wang, H., 2016. Correlation Analysis Using Global Dataset of Events, Location and Tone. 2016 IEEE First International Conference on Data Science in Cyberspace (DSC), pp. 648-652.

Colesky, M., Hoepman, J.-H. & Hillen, C., 2016. A critical analysis of privacy design strategies”, Procs. IWPE’16, IEEE, 33–40.. 2016 IEEE Security and Privacy Workshops (SPW), pp. 33-40.

Glasgow, K., 2015. Big data and law enforcement: Advances, implications, and lessons from an active shooter case study. In: Application of Big Data for National Security. Waltham: Butterworth-Heinemann, pp. 39-54.

Glassman, M. & Kang, M. J., 2012. Intelligence in the internet age: the emergence and evolution of OSINT. Computers in Human Behavior, Volume 28, pp. 673-682.

Guest Editorial, 2013. Legal aspects of open source intelligence - Results of the VIRTUOSO project. Computer law & security review, Volume 29, pp. 642-653.

Guo, L. & Vargo, C., 2017. Global Intermedia Agenda Setting: A Big Data Analysis of International News Flow. Journal of Communication , pp. 499-520.

Hribar, G., Podbregar, I. & Ivanusa, T., 2014. OSINT: A ‘‘Grey Zone’’?. International Journal of Intelligence and CounterIntelligence, Volume 27, p. 529–549.

Hustinx, P., 2010. Privacy by design: delivering the promises. Identity in the Information Society, 3(2), pp. 253-255.

Jain, P., Gyanchandani, M. & Khare, N., 2016. Big data privacy: a technological perspective and review. Journal of Big Data.

Koops, B., 2013. Police investigations in Internet open sources: procedural law issues. Computer Law & Security Review, Volume 29, pp. 676-688.

Koops, B., Hoepman, J. & Leenes, R., 2013. Open-source intelligence and privacy by design. Computer Law & Security Review, Volume 29, pp. 676-688.

Krempel, E. & Beyerer, J., 2014. TAM-VS: A Technology Acceptance Model for Video Surveillance. In: Privacy Technologies and Policy. Cham: Springer, pp. 86-100.

Kung, A., 2014. PEARs: Privacy Enhancing ARchitectures. In: Privacy Technologies and Policy. Cham: Springer, pp. 18-29.

MARISA, 2018. D3.2 MARISA SERVICES DESCRIPTION DOCUMENT, s.l.: s.n.

Martín, Y.-S. & Kung, A., 2018. Methods and Tools for GDPR Compliance through Privacy and Data Protection Engineering. 2018 IEEE European Symposium on Security and Privacy Workshops, pp. 108-111.

Matturdi, B., Zhou, X., Li, S. & Lin, F., 2014. Big Data security and privacy: A review. China Communications, pp. 135-145.

Moallem, A., 2019. Perspectives on the future of human factors in cybersecurity. In: Human-computer interaction and cybersecurity handbook. Boca Ratom: CRC Press, pp. 353-366.

Nelson, B. & Olovsson, T., 2016. Security and privacy for big data: A systematic literature review. 2016 IEEE International Conference on Big Data (Big Data), pp. 3693-3702.

Rahman, F., 2017. Smart Security: Balancing Effectiveness and Ethics. RSIS Commentary, 14 Dec.Volume 235.

Rumbold, B. & Wilson, J., 2018. Privacy Rights and Public Information. The Journal of Political Philosophy.

Staniforth, A., 2016. Big Data and open source intelligence – A game-changer for counter-terrorism?. TRENDS Research & Advisory, 19 July.

Van Aubel, P. et al., 2018. Privacy by design for local energy communities. Ljubljana, s.n.

Wójtowicz, A. & Cellary, W., 2019. New challenges for user privacy in cyberspace. In: Human-computer interaction and cybersecurity handbook. Boca Raton: Taylor & Francis Group, pp. 77-96.

Yin, R. K., 2009. Case Study Research Design and Methods. s.l.:Thousand Oaks: Sage Publications.

Zielinski, A., 2013. Detecting natural disaster events on twitter across languages. Intelligent interactive multimedia systems and services. 6th International onference on Intelligent Interactive Multimedia Systems and Services, pp. 291-301.

Jyri Rajamäki is Principal Lecturer in Information Technology at Laurea University of Applied Sciences and Adjunct Professor of Critical Infrastructure Protection and Cyber Security at University of Jyväskylä, Finland. He holds D.Sc. degrees in electrical and communications engineering from Helsinki University of Technology, and PhD degree in mathematical information technology from University of Jyväskylä.

Dr Trishana Ramluckan a Post- Doctoral Researcher in International Cyber Law, College of Law and Management Studies, UKZN. Her areas of research include IT and Governance. She is a member of the IFIP working group on ICT Uses in Peace and War and is an Academic Advocate for ISACA.

Juhani Rauhala is a Research Affiliate and PhD student of cybersecurity at the University of Jyväskylä. He has over ten years' experience in the telecommunications industry and has been awarded two patents related to cloud storage. Juhani is a designated Eur Ing by the Federation of European Engineers and holds BScEE (1992) and an MScE (1996) degrees from San Francisco State University. His research interests include the weaponization of ubiquitous technologies and technology abuse.

Dr. Aunshul Rege is an Associate Professor with the Department of Criminal Justice at Temple University. Her cybercrime/security research on adversarial decision-making and adaptation, organizational and operational dynamics, and proactive cybersecurity is funded by several National Science Foundation grants.

Dr. Mari Ristolainen is a Researcher at the Finnish Defence Research Agency. She has studied psychology at the Moscow State University and she earned a doctorate in Russian Language and Cultural Studies from the University of Joensuu in 2008. She has been conducting postdoctoral research in the field of Russian and Border Studies in several Academy of Finland- and EU-funded projects at the University of Eastern Finland and at the University of Tromso, Norway. Her current research interests include cyber warfare as a phenomenon, Russian digital sovereignty, and the governance of cyber/information space.

M.Sc. (Cognitive Science) Tarja Rusi is a Cyber Security Master's student (University of Jyväskylä, Finland). She has 20+ years career in the telecommunications industry and has been lecturing on cyber threats and cyber terrorism. She has participated in governmental cyber threat evaluation work. Research interests: critical infrastructure protection, cyber threats, cyber terrorism, state sponsored cyber-attacks.

Helvi Salminen has worked in information security since June 1990. Before her security career 12 years of experience in systems development. Helvi is founder member of Finnish Information Security Association and president of ISACA Finland Chapter Helvi is qualified CISA, CISSP & SABSA & was awarded as CISO of the year in Finland 2014.

Dr. Char Sample is research fellow for ICF Inc. at the US Army Research Laboratory in Maryland, and is a visiting academic at the University of Warwick, UK. She has over 20 years experience in the information security industry and focuses her research on Fake News, cultural values in cyber security events, and data resilience.

Leonel Santos is an Equiv. Assistant Professor at Polytechnic Institute of Leiria. He is a researcher on the Computer Science and Communication Research Centre and is a PhD student in University of Trás-os-Montes e Alto Douro. His major research interests include Cybersecurity, Information and Networks Security, Internet of Things, Intrusion Detection Systems and Computer Forensics.

Mr. Lynn Scheinman is a technology consultant specialising in defence, security, and applied data science at SAP in Walldorf, Germany. He received his Masters in Science of Project Management and Operations Research at Florida Institute of Technology. His defence experience comes from 10 years in the US Army Special Forces as a Green Beret.

Dr Keith Scott is the Subject Leader for Languages at De Montfort University, where he is also a member of the Cyber Security Centre. He is also a member of the Cyber Policy Centre, a UK-based independent public policy centre devoted exclusively to the consideration of cyber as a socio-technical phenomenon. His research interests include the social and cultural implications of 'cyber' as a concept, influence, online communication, and the use of gaming as a teaching and research tool.

Youngsup Shin is a researcher in Agency for Defense Development, South Korea. He is in an integrated PhD program in Korea University. His main research areas are cyber situational awareness and cyber warfare.

Ph.D. Petteri Simola is a senior psychologist at the Finnish Defence Research Agency, Human Performance Division. His work involves human aspects of information security, Human Factors (especially sleep) and aptitude testing in recruitment.

Jussi Simola is a PhD student of cyber security in University of Jyväskylä. His area of expertise includes decision support technologies, SA systems, information security and continuity management. His current research is focused on effects of cyber domain as a part of Hybrid Emergency Response Model.

Mr Veikko Siukonen is a research officer at Finnish Defence Research Agency (FDRA). He resieved his Master's Degree in Military Sciences from The National Defence University in 2007. He is a master of science student (cyber security program) in University of Jyvaskyä. His main research areas are cyber warfare and cyber threat intelligence.

Tiia Sõmer is early stage researcher at Tallinn University of Technology. She is conducting PhD level studies, focusing on modelling cyber criminal journey mapping. In addition to research she does teaching and has co-authored educational materials for general education. Before starting academic career, she served for more than twenty years in the Estonian defence forces.

Lee Speakman, Lee gained his PhD in the area of Mobile Ad hoc Networks from Niigata, Japan, in 2009. Since then Lee worked in the area of networks, network security, and software exploitation and protection measures in Defence. Lee joined the University of Chester in 2015 to develop and deliver the University's new Cybersecurity programmes and research.

Ilona Stadnik is a PhD student at the School of International Relations, Saint-Petersburg State University, Russia. During 2018-2019 academic year she was a Fulbright visiting researcher at Georgia Institute of Technology, USA, working with Internet Governance Project. She has been a regular participant and speaker at major cybersecurity events such as the United Nations Internet Governance Forum (IGF), CyFy conference, European Dialogue on Internet Governance (EuroDIG). Her research covers international cyber norm-making, Russia-US relations in cybersecurity, and global Internet governance.

Dr. Nikolai Stoianov is Colonel in the Bulgarian armed forces, Deputy Director of the Bulgarian Defence Institute "Prof. Tsvetan Lazarov" and principal member of NATO's Science and Technology Board. He is also associate professor and leads several international research projects on cybersecurity and related issues.

Mr Marcel Stolz is a doctoral student in cyber security at the University of Oxford, UK. He has a background in Computer Science and has served as a First Lieutenant in the Swiss Armed Forces. His research interests lie in global cyber security and regulation of data companies, such as Facebook.

Dr. Steven Templeton is a researcher at the University of California, Davis, USA. Since 1999 he has operated a consulting firm specializing in ICS security and compliance. Originally a wildlife biologist, in 2018 he received his PhD in computer science. His research spans multiple area of computer security, in particular intrusion detection, monitoring, and attack modelling.

Dr Ben Turnbull is a Senior Lecturer for the University of New South Wales, Australia. He is an expert in cyber security and digital forensics with 16 years in the industry. He is also a Certified Information Systems Security Professional (CISSP). He has previously worked as a research scientist for the Defence Science and Technology Organisation in the field of cyber network defence and analysis. In his spare time, Ben plays too many card and board games.

Maija Turunen is a PhD Student at the Finnish National Defense University. Her main research areas consist of cyber warfare, Russia and strategic communication. Maija Turunen works as a legal counsel at the Finnish Transport Infrastructure Agency.