The Right to Erasure and its Implication on AAL Systems

Kristin Aleksandrova

Sofia University St. Kliment Ohridski, Sofia, Bulgaria

kristinia@uni-sofia.bg

Abstract. With the extended life expectancy, we have seen an increase in the load put on each country's healthcare system. This has increased funding for technologies that could enable the autonomous living of elderly or disabled people. In general, those technologies can be condensed under the premise of Ambient Assisted Living (AAL) Systems. These systems aim to improve the quality of life, by utilizing a multitude of technologies – mainly assistive technologies, parts of smart home solutions and telehealth services. As research has been funded by organizations with a specific use case in mind, usually an elder care facility or a caretaker organization, the topic of data privacy has been widely overlooked. In principle, there is a trade-off between the functionality and the affected person even in data privacy regulations there is an exception for these types of healthcare cases. This however closes the door for machine learning enhancements of AAL systems. Most of the current approaches rely on processing the data in real time and not persisting it; this creates a semblance of security, but provides no approach for the problem resolution. To enhance a system with a data model of any kind we would need to store and analyze the data in compliance with the current data privacy regulations. For the purpose of this work, we focus on the General Data Protection Regulation (GDPR), and the implications of the Right to erasure specify on AAL systems.

Keywords: Right to Erasure, Elderly Care, AAL Systems.

1 Introduction

The mean age of the world's population is rapidly increasing, due to a variety of factors, one of which is the longer life expectancy, to an extent due to significant improvements in our healthcare. The current prognosis suggests that by 2050 the aging population will double in size compared to 2017, a staggering 2.1 billion against the current 1 billion. This is projected to put an additional load on each country's healthcare system. As we can see by the graphic provided by Our World in Data (fig. 1) for the past 50 years, there has been an increase in the investment in healthcare. While there is a multitude of factors, such as modernization and medicine costs, the trend is that this investment would continue to increase and one of the major factors contributing would be supporting the elderly.

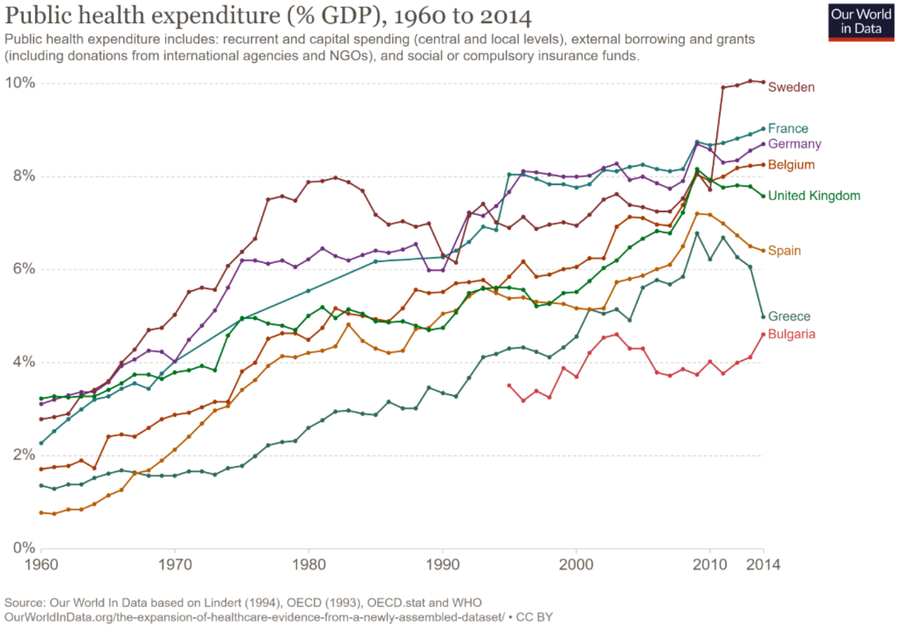

Fig. 1. Public health expenditure as percentage of GDP for European countries, based on data from Lindert (1994), OECD (1993), OECD. Stat and WHO, visualized by Our World in Data.

To understand what is causing this economic load and why is it such a topic of discussion in recent years, let us look at the work of Sarah Abdi, Alice Spann, and Jacinta Borilovic [1]. They define three categories in which elderly people need support: social activities and relationships; psychological health; activities related to mobility, self-care, and domestic life. To address that need we see a rise in Ambient Assisted Living (AAL) systems, as their purpose is to reduce the load on formal health and care structures, while providing independence for the senior population. Depending on their functionality and technology, we can separate these systems in distinct categories. One such differentiation, based on the primary function of the examined AAL systems, is done in the taxonomy created by Byrne et al. [5]. Here we see four major classes: Smart Homes; Intelligent Life Assistants; Wearables; Robotic Assistance. It is obvious that these classes have a different target user group and underlying technology.

One commonality, of interest for us, is the handling of data privacy. Most systems rely on the crucially of the provided functionalities, considering these systems can significantly improve a person's independence and daily life, in addition to providing crucial information in life-threatening situations. Data is collected via a multitude of channels, including mobile phones and is processed real-time with no persistence. This however limits the ability of these systems to grow by further analyzing historical data and behaviors and being proactive instead of reactive.

1.1 Personal data in AAL systems, subject to GDPR

Based on their primary function, different AAL systems store diverse types of personal data, most of it sensitive. An important thing to note is the difference between personal and sensitive data. Sensitive data is a subset of personal data, in more details – personal data is considered sensitive if it is personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data processed solely to identify a human being; health-related data; data concerning a person's sex life or sexual orientation. Sensitive data is subject of additional processing conditions, for example according to GDPR sensitive data should be held separately from other personal data. We would not go in the details of that segregation, but sensitive data will be interchanged with personal data when speaking about heal-related data in this paper.

In principle, all Ambient Assisted Living systems store information about the individual they are assisting and at least one caretaker, responsible for them. This role can be taken by a family member, friend, or a medical professional. Therefore, the basic personal data of those two roles is stored as well as their relationship: 'C' is the caretaker of 'I' and as such is using the AAL system and accessing some of the sensitive data of 'I'. This basic information consists of some form of identification, such as a full name, and some means of contact, such as a phone number, email, or physical address.

Based on the system functionality, the system can store GPS location data, relating to the individual's location, collected real-time; medical information, such as a person's diagnosis, medical prescription, test results, etc.; information for their daily schedule and habits, including places they often visit; additional streams of information from wearable or IoT sensors [15]. This data can be used to derive additional personal information for an individual with malicious intent. Considering the main target of AAL systems are the elderly, detailed information about their family members can be abused for extortion. Alternatively, data about their schedule and location can easily facilitate theft and other crimes.

1.2 Right of Erasure

In 2018, the European Union passed the General Data Protection Regulation (GDPR). One of its aspects is the so called "right of erasure". It introduces the right a person has, to request verbally or in writing, the erasure of personal data of his that is being collected. Of course, there are exceptions. For example, if the processing is necessary for public health purposes in the public interest.

Additionally, if the processing is necessary for the purposes of preventative or occupational medicine, for the working capacity of an employee, for medical diagnosis, for the provision of health or social care or for the management of health or social care systems or service [2]. From this, a natural conclusion is that AAL systems can be categorized as cases where an exception is to be made concerning the “right of erasure”. While this is a positive factor for the fast functional growth of AAL systems, it is hindering their wider adoption. Their usage at present is limited to cases where the data processing can be correlated with definitive medical benefits, considering the diversity of functionalities and the problem areas AAL systems cover, this correlation is hard to illustrate. To utilize these advancements for the public, regardless of their condition, GDPR and the “right of erasure” need to be enabled.

Considering GDPR came to force three years ago, it has already been implemented in many software solutions. Let us take for example, how the right of erasure affected the way Google was displaying results, naturally a simple search could provide a decent collection of personal data for a target individual, based on each website’s policy. The first changes were implemented in 2014, when it was made possible to request the exemption of certain URLs, containing personal or sensitive data from the search results in Europe. In principle, if an individual finds personal or sensitive data of theirs in a search result, they can request that the result be removed. Google does not remove or restrict any public site, instead they de-list them, as also seen their erasure request form [3]. In addition, they are legally obliged to do so only for search results on the territory of Europe [4]. In turn, this poses a new perspective on the legislation. You do not need to physically remove the data, provided you as processor do not use, expose, or make it possible for a third party to find the data in question.

2 Implications for standard AAL systems

To understand the implications of the right to erasure on AAL systems, we need to consider the data that is being collected. Most systems have a specific purpose and use-case for which they are developed, for example, there are systems that aim to help dementia patients and mitigate their wandering behavior, and alternatively there are fall-detection systems that prioritize the fast recognition and timely reaction in case an elderly person has fallen. There are many more types of systems [5]. Looking at the problem areas they are addressing we already can anticipate the difference in the used technologies and therefore the data collected.

Let us take as an example Wandering Prevention Tools (WPT), which are addressing the wandering behavior of wandering patients. To do that, these systems collect and analyze real-time location data, to recognize a wandering episode. Afterwards a responsible person needs to be notified, so that adequate actions are taken. This leads to data being centrally stored or at least centrally processed and analyzed. This in turn complicates the insurance of data privacy and the right of erasure, as upon request, related data from a multitude of devices needs to be deleted. In one such system that addresses nighttime wandering behavior [6], we can see that data is gathered from multiple sensors and devices and persisted in an OpenRemote server [7], afterwards the taken actions consist of changes in the person's environment to stimulate their return to bed. Looking at the architecture of OpenRemote and their privacy policy [8], we can conclude that sensor data is persisted in the local controller. Considering the nature of the action taken never leaves the premise of the home system. In this case, a request to delete all collected personal data is locally executed and the system will lose all previous information about a person. On the other hand, systems like Carelink [9], which is a project funded by the EU's AAL Program, rely on data collected from tag sensors and analyzed in a cloud-based platform. This aims to enable family and caretakers to receive up-to-date information about an individual. Besides the increased need for data privacy insurance, a request for data erasure would mean to remove of location-based data, all information about the individual. In addition, devices he is identifiable by, all information about family members and relational data, that has been gathered and in case this is the only person in the system, a caretaker is responsible for, the caretaker's data as well.

Besides the purely technical limitations for data removal, there is one additional aspect that is best seen in AAL systems based on smart home/IoT solutions. The right to erasure ensures that any individual can request from the data controller the removal of all collected personal data. In the case of an AAL system, that relies on a variety of sensors and smart home appliances from different manufactures with different communication concepts; a question arises about who is the data controller. Each device can send data to its own cloud location, independently from the AAL system processing, regardless that data may be included in the erasure request. This means that not only we need to establish that personal data should be easily removable from the system at a low cost, but we need to ensure that there is a responsible party for the removal of said data. In an article [10] in the International Data Privacy Law journal, the concept of joint controllership and its implications is explored in detail. The basic premise being that this responsibility needs to be shared. For AAL systems, they are the obvious choice as a data controller; however, that implies that they need to have control over the data processing by any "third-party", which includes devices and sensors part of that system. Usually that is not the case, these sensors and devices are owned by separate companies and without additional consent, and the AAL system cannot represent the individual in their request for erasure. Not resolving that legal complication can render a GDPR-enabled system, noncompliant with the European legislation.

3 Implications for AAL systems with ML functionality

One emerging area of extension for AAL systems is the addition of "smart insights", or in other words features that aim to give proactive signals and knowledge in the respective use-case of the system. This usually involves the application of one of the well-known machine learning algorithms to the gathered data. Naturally, this carries additional implications for data privacy and especially the right of erasure. An extensive research on the topic has been published in the Computer Security & Law Review journal [11]. A question, it aims to answer, of relevance to us is if a data model has been trained using personal data, what does it mean to erase an individual's data? We need to delete the data that is stored for training purposes. However, do we need to remove it from the model, or in other words retrain it? If not, how do we ensure continuity between the different versions of the model as each time the training data is completely different. At present, there is no clear strategy from the EU on how to ensure GDPR compliance in a machine-learning context. Each data controller settles on a compromise, as the current GDPR regulation and the standard machine learning development are not compatible.

If we look at the initial example with Google's approach to the right of erasure, we need to ensure that in the AAL system there is no identifiable data. In which case we would not be required to delete it. Let us consider a few of the more common approaches. Pseudo anonymization replaces the personally identifiable information with artificial identifiers, so that the data cannot be linked to an individual with additional information that is stored separately. However, many pseudo anonymization algorithms are reversible in which case they are not exempt from the right to erasure [12]. If we anonymize all data, we are by design compliant with GDPR, as there is no identifiable personal data, processed by the AAL system. However, we lose the connections different data sources have, as they describe the same person's behavior or the same event that has been recorded. In many cases, it has been proven that the quality of a machine-learning model significantly declines when the training data is anonymized [11]. For completeness, we should also mention cryptographical functions; they establish an isomorphism, which allows us to do mathematical operations on the encrypted data without deciphering it. Considering the amount of raw data and noise AAL systems generate daily, this does not appear as a preferable approach since in practice the algorithms known today are inefficient for large data sets.

For example, Valenzuela et al. [14] proposes an intelligent system that can be integrated into a wearable device, and is able to diagnose cardiac diseases in real time. This done with the assistance of a genetic algorithm, trained on seven different electrocardiogram datasets. If we assume such a wearable heart monitor device is part of an Ambient Assisted Living system, with the purpose of timely diagnosis of cardiovascular diseases, then the system itself can maintain a dataset of electrocardiograms of its users. The main benefit of training a model on the data gathered by the concrete devices, part of the AAL system, instead of the generic datasets, is that the training data would be close to the real-world data it would be applied. There will be the same device specifics and outlier behaviors. The problem comes, when one of the monitored individuals invokes the right to erasure. For the AAL system to keep using the provided heart monitor data, we need to ensure that it is untraceable to its origin. Here we hit all the concerns, previously described around anonymization. Alternatively, if we were to remove the records from the training dataset, we could be removing crucial to the algorithm records. In turn we have no way of predicting the immediate impact of other users of the AAL systems and the behavior of the model if this should be reversed and the requester decides to renew using the AAL system and its functionalities.

This leaves us with a few options, that deciding what to compromise with. We can lose functionality or certain system capabilities by pre-processing the data in such a way that we are exempt from the right to erasure or we can look for a design that allows us to work with the full set of data and relationships, in compliance with GDPR.

4 Discussion and conclusions

In the previous sections, we saw, that in the current implementations of AAL systems, there are different approaches for data handling. Some systems analyze data real-time, while others persist the data and provide long-term analytics and statistics. Even the latter ones are divided based on the location, where said aggregation is occurring. For most systems, a centralized cloud-based solution receives all data and performs the analysis. The other approach works with the data in a local isolated environment. The case is with the wandering prevention tool that we took as an example. In that case, the architecture of the AAL was based on OpenRemote and this in turn ensured that personal data does not leave the premise of a person's home. If we restrict the data processing and persistence to a local server, we minimize the risk of unauthorized data exposure and we simplify the compliance with GDPR. This approach is clearly most easily applicable for AAL systems that are classified as Smart Home solutions. To summarize, we can argue that the right to erasure is most easily implemented in AAL systems that are based on Smart Home solutions, as the data processing is easily restricted for localized processing.

There has been a rise in the applications of machine learning technologies is AAL systems. However, as the field is relatively new, there is a more prominent focus on functionality instead of data privacy. This is possible due to the available exception cases in GDPR and similar privacy legislations that allow data privacy to be overlooked in the interest of public health and the provisioning of crucial life-sustaining services. Nevertheless, there is an opportunity for newly developed or extended AAL systems to implement a Privacy by Design [13] approach. If we combine this with the general implications that GDPR and the right to erasure have on AAL systems, we can argue that a personalized model, trained in a smart home environment is the easiest intelligent extension of AAL systems. Similarly, to the standard case, where data is more secure and easily mailable as per GDPR, when it is locally analyzed – a data model that has used sensitive data as training data would be easier to comply with the right to erasure if it is personal and localized to a home environment. We saw that anonymization or pseudo anonymization only partially resolves the issue and we still face a decline in the trained model quality. The same would happen if we have a model trained on data provided by different patients. We have no empiric measures for most algorithms to explain what the weight of a certain person's data is in the overall model. Therefore, we also cannot outline the decrease in a model's performance if we were to remove an individual and their data. A downside would be the slower ramp-up time, during which data is collected and analyzed, without that phase we will have no insights whatsoever that can be provided. Nevertheless, this downside is easily mitigated by the non-existent implications of removing an individual and their data from the trained model, as the removal itself would mean a deletion of their data and models on a local level, usually the smart home middleware. In that case, explaining the implication of evoking the right to erasure is also simplified – the AAL system will return to its initial state and for it to be able to provide these types of insights again a similar in duration training period needs to occur.

This hypothesis is based on the desire to comply with the European GDPR legislation and specifically the right to erasure and its implications. Regarding data privacy there are many additional aspects that need to be considered for an AAL system to be fully compliant with GDPR. Those would be the focus of a future work. In addition, the proposal for a personalized model, locally managed via a middleware, needs to be supported by those additional aspects of GDPR. Overall, the end-goal of this confirmation would be the creation of a prototype solution, illustrating the benefits of a data privacy aware AAL system.

References

1. Abdi, S., Spann, A., Borilovic, J. et al. Understanding the care and support needs of older people: a scoping review and categorization using the WHO international classification of functioning, disability and health framework (ICF). BMC Geriatr 19, 195 (2019).
2. Right to Erasure as described by the Information Commissioner's Office, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/, last accessed 2021/04/18
3. Google's EU Removal form, https://www.google.com/webmasters/tools/legal-removal-

request?complaint_type=rtbf&visit_id=637202230061146146-20083139&rd=1, last accessed 2021/04/18

4. BBC article “Google wins landmark right to be forgotten case”, https://www.bbc.com/news/technology-49808208, last accessed 2021/04/18
5. Byrne, Caroline & Collier, Rem & O’Hare, Gregory. (2018). A Review and Classification of Assisted Living Systems. Information.
6. Radziszewski, Robert & Ngankam, Hubert & Pigot, Hélène & Grégoire, Vincent & Lorrain, Dominique & Giroux, Sylvain. (2016). An ambient assisted living nighttime wandering system for elderly. 368-374
7. OpenRemote main site, https://openremote.io/, last accessed 2021/04/18
8. OpenRemote privacy policy, https://openremote.io/privacy-policy/, last accessed 2021/04/18
9. CareLink main site, http://carelink-aal.org/, last accessed 2021/04/18
10. Jiahong Chen, Lilian Edwards, Lachlan Urquhart, Derek McAuley, Who is responsible for data processing in smart homes? Reconsidering joint controllership and the household exemption, International Data Privacy Law, Volume 10, Issue 4, November 2020, Pages 279–293
11. Fosch Villaronga, Eduard and Kieseberg, Peter and Li, Tiffany, Humans Forget, Machines Remember: Artificial Intelligence and the Right to Be Forgotten (August 13, 2017). Computer Security & Law Review (Forthcoming),
12. Hintze M., El Emam, K.: Comparing the Benefits of Pseudonymization and Anonymization Under the GDPR (17th August 2017)
13. Sylvia Kingsmill, Dr. Ann Cavoukian,: Privacy by DesignSetting a new standard for privacy certification, https://www2.del
14. oitte.com/content/dam/Deloitte/ca/Documents/risk/ca-en-ers-privacy-by-design-brochure.PDF, last accessed 2021/04/18
15. Valenzuela O, Prieto B, Delgado-Marquez E, Pomares H, Rojas I (2018) Wearable intelligent system for the diagnosis of cardiac diseases working in real time and with low energy cost. Multidiscip Digit Publ Inst Proc 2(19):513
16. Mulvenna, Maurice & Carswell, William & Mccullagh, Paul & Augusto Wrede, Juan & Zheng, Huiru & Jeffers, W. & Wang, Haiying & Martin, Suzanne. (2011). Visualization of Data for Ambient Assisted Living Services. IEEE Communications Magazine. 49. 110-117. 10.1109/MCOM.2011.5681023.