Abstract

**Background:** Personal Health Data (PHD) research has been intensified over the last years, attracting the attention of scientists from different fields, such as Software Engineers, Computer Scientists and Medical Professionals. The increasing interest of researchers can be attributed to the exponential growth of the available PHD due to the widespread adoption of ubiquitous technology in everyday life, as well as to the potential of the ongoing digital transformation in healthcare. This increasing interest requires that academia has an overview of the published scientific literature to plan future endeavors.

**Objective:** The main objective of this study is to identify and address research gaps in literature regarding PHD.

**Method:** This paper conducts a systematic mapping study to summarize the existing PHD approaches in literature and to organize the selected studies according to six classification criteria: publication source, publication year, research types, empirical types, contribution types and research topic.

**Results:** In total 79 papers have been included after fulfilling the inclusion...
criteria and have been classified accordingly. There is an increasing amount of attention that has been paid to PHD since 2014. The majority of papers is published in journals. The two main research types found were solution proposals and evaluation research. The majority of the selected papers were empirically evaluated. The main contribution types were methods and frameworks. Data privacy is the most frequently addressed topic in PHD literature, followed by data sharing.

Conclusions: The findings of this systematic mapping study have implications for both researchers who are planning new studies in PHD and for practitioners who are working in connected health and would like to have an overview on the existent studies on PHD research area.

Keywords: Personal health data, systematic mapping study, digital data, data privacy and security, data sharing

1. Introduction

A demographic change has been projected worldwide, the populations are rapidly aging. According to World Health Organization (WHO), “Between 2015 and 2050, the proportion of the world’s population over 60 years will nearly double from 12%, to 22%. By 2020, the number of people aged 60 years and older will outnumber children younger than 5 years” [1]. This shift in the distribution of the global population towards aging poses significant challenges to countries as they have to ensure that social and healthcare systems are aligned with the needs of older populations. Even so, as people grow older tend to develop more complex health conditions increasing the likelihood of
having one or multiple chronic diseases. Traditional healthcare services are often designed to manage health conditions in disconnected and fragmented ways so that in many cases hinder the communication between patients and healthcare actors. To support this change future healthcare systems need to undergo a transformation so that can ensure effortless communication and coordination between stakeholders. Advances in digital health could shift the tide of healthcare services towards long-term and sustainable solutions.

Technological innovations in digital health have been seen to hold the potential for reformation of healthcare services. The development of information and communication technologies (ICT) and their adoption in healthcare industry is changing the provision of services, not only improving the amount and quality of them, but also facilitating effective communication between patients and medical professionals [2]. It has been foreseen that ICT will facilitate the shifting center of care from physicians to patients and to self-management of diseases. Current digital solutions such as home nursing [3], health monitoring systems [4], mobile applications [5], electronic medical records (EMR) [6], electronic health records (EHR) [7], personal health records (PHRs) [8] and mobile PHRs (mPHRs) [9], can support this argument.

In the same vein, EHRs’ format standardization and expansion of access to patients is a milestone towards the promotion of self-management of diseases [8, 10]. By granting patients with access and ownership to EHR, a primary form of PHRs emerged by EHR [11]. The context of EHR is restricted to medical information and it is useful for diagnosis of diseases [12, 13], so that patients have limited contribution and control over them
In contrast to EHRs, PHRs hold health and wellness data that are exclusively controlled by patients across their life span. Besides that, PHRs can support patients to make health decisions and to self-manage their health [8, 14]. The first discussions about PHRs were held in late 1970’s [15]. The original idea emerged from the need to personalize technologies and to make health records accessible to the general population. PHRs can facilitate patients to self-monitor their own health and enable them to store information about behaviors, medical treatments or allergies [16]. PHRs can stand-alone on personal computers [17] or can be a part of a system administered by healthcare organizations [8]. Besides PHRs and EHRs, mPHRs allow users to access and coordinate their life long health information through their smartphones and enable them to distribute their personal health data (PHD) when necessary [9, 18]. The system combining EHR and mPHR is considered a powerful tool that can facilitate seamless communication between patients and medical professionals leading to faster and informed decisions, especially in cases of patients with chronic conditions [19]. Unified systems combining PHRs and EHRs could maximize patients’ engagement and increase their participation to shared decision-making processes [20].

Towards the same direction, the proliferation of network sensors either worn or placed in the living spaces and connected to the Internet of Things (IoT) could bring patients one step closer to personalized healthcare [21, 22]. Wearable sensors, for example smartwatches or accelerometers and gyroscopes built-in smartphones, allow real-time physiological monitoring, providing thus PHD about daily habits in real-world contexts [23]. PHD has the potential to revolutionize the landscape of healthcare services providing
information that could lead to customization of medical treatments and even to diseases prediction leveraging big data classification techniques [24]. Big data analytics technology holds the potential to transform health information by harnessing complex and large amount of datasets to infer real-time knowledge and new insights from it [25, 26, 27].

PHD is a research topic that has attracted the interest of researchers and practitioners from various disciplines, such as Computer Scientists, Engineers and Medical Professionals [28, 29, 30]. However, to the best of our knowledge, no previous study has conducted a systematic mapping study on PHD. The aim of this paper is to identify and map the existence research in PHD research area, addressing possible research gaps and suggesting future research directions. In this paper, a systematic mapping study has been performed, to answer six mapping questions (MQs), resulting to the selection and classification of 79 papers. The classification criteria are the following: publication source, publication year, research types, empirical types, contribution types and research topic.

The remainder of paper is organized as follows. Section 2 provides a summary of related work concerning PHD. Section 3 outlines the research methodology of this paper. Section 4 reports the classification results, while Section 5 discusses the principle findings, presents implications for researchers and practitioners and addresses the limitations of this study. Finally, Section 6 presents the conclusions and proposes work that will be carried out in the future.
2. Background and significance

Various aspects of PHD have been studied and discussed in literature. For instance, the value of visualization modalities [28], mobile-based solutions for integration of PHD to healthcare systems [31] and strategies for development of infrastructures for management of PHD [32]. Many researchers have carried out studies focusing on security and privacy issues of PHD proposing solutions for minimizing data breaches [33], or legal frameworks, policies and ethical considerations for protection of PHD [34, 35, 36]. PHD management leveraging meta-data and cloud-computing for seamless access and sharing of data among medical professionals [37], as well as access control and security of PHD in cloud-computing have also been the subject of discussions [38]. Moreover, the potential of data mining and big data techniques in healthcare have been reviewed [39, 40]. Studies have also addressed challenges concerning sensor interoperability in terms of infrastructure capabilities [41] and data fusion conflicts for the integration of sensor data to EHR [42]. Previous studies concluded that patients with long-term conditions, such as diabetes and congestive heart failure, or medical treatments as vitro fertilization, which creates new needs for information or communication, are the users that will probably visit their health records more frequently [43, 44, 45].

Over the last years researchers have been given attention to PHD, for this reason we believe that is important to present an overview of this research area. Systematic literature reviews (SLR) and systematic mapping studies have been used to give an overview of the state-of the art of different disciplines [46]. The benefits and drawbacks of conducting literature reviews have been addressed by various studies [47, 48]. In this paper, a system-
atic mapping study has been used to provide an overview of the available papers. Our aim is to classify the selected studies, based on the established classification scheme. To the best of our knowledge, no previous studies have reported and mapped PHD using this classification scheme. Nevertheless, Acher et al. [17] and Roehrs et al. [49] have conducted literature reviews on PHRs. They have presented previous literature using different classification schemes and sets of research questions. Besides that, Gagnon et al. [50] have presented a contribution about electronic PHR (ePHR) reporting federal and local laws and policies that influenced ePHR implementation in Canada. Moreover, a literature review focusing on consumer perspectives of PHR has been conducted [51].

3. Methodology

The systematic mapping study principal goal is to provide an overview of a research area, and identify the quantity and type of research and results available within it. A mapping process consists of three activities: the search for relevant publications, the definition of a classification scheme and the mapping of publications [52]. This method focuses on classification, conducting a thematic analysis and identifying publication fora. Fig. 1 shows the mapping process, which covers the search for relevant publications, the definition of a classification scheme, and the mapping of publications.

3.1. Mapping questions

This study aims to gain insight into the existing literature about PHD. The systematic mapping study therefore addresses six MQs. The MQs with the rationale motivating the importance of these questions are presented in
Table 1: Mapping questions

<table>
<thead>
<tr>
<th>ID</th>
<th>Mapping question</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQ1</td>
<td>Which publication channels are the main targets for PHD research?</td>
<td>To identify where PHD research can be found as well as the good targets for publication of future studies.</td>
</tr>
<tr>
<td>MQ2</td>
<td>How has the frequency of studies related to PHD changed over time?</td>
<td>To identify the publication trends over time of PHD literature.</td>
</tr>
<tr>
<td>MQ3</td>
<td>What are the research types of PHD studies?</td>
<td>To explore the different types of research reported in the literature concerning PHD.</td>
</tr>
<tr>
<td>MQ4</td>
<td>Are PHD studies empirically validated?</td>
<td>To discover whether research on PHD has been validated through empirical studies.</td>
</tr>
<tr>
<td>MQ5</td>
<td>What are the approaches that were reported in PHD research?</td>
<td>To discover the existing PHD approaches reported in the existing PHD literature.</td>
</tr>
<tr>
<td>MQ6</td>
<td>What are the main topics in PHD literature?</td>
<td>To identify the research areas related to PHD.</td>
</tr>
</tbody>
</table>

Table 1. The search strategy and paper selection criteria were defined on the basis of them.

3.2. Search strategy

The papers were identified by consulting the following sources: IEEE Digital Library, ACM Digital Library, Science Direct and SpringerLink. Google scholar was also used to seek literature in the field. The search was performed in January 2018. The search string used to perform the automatic research in the digital libraries selected was formulated as follows: “Personal” AND
“health” AND “data”. This search string was applied in the title, abstract and keywords of the investigated papers to reduce the search results. The search string was limited to three keywords as the authors’ search strategy was to include a broad selection of relevant literature in the first iteration, relying afterwards on the screening of papers to include or exclude them in the study [46].

3.3. Paper selection criteria

Each paper was retrieved by the first author and the information about it was filed in an excel file. All results from the search engines were retrieved and analysed. The author considered each paper’s title, abstract and keywords, and commented on whether the paper should be included or excluded in the excel file. The second author revised the final selection. The first step after the application of the search string was to eliminate duplicate titles, and titles clearly not related to the review. The inclusion criteria (IC) were limited to:

IC The studies that address personal health data.

The studies that met at least one of the following exclusion criteria (EC) were excluded:

EC1 Papers that focus on personal pets.
EC2 Papers whose subject is non-digital health data.
EC3 Papers that focus on systems for data collection.
EC4 Papers that focus on information systems interoperability.
**EC5** papers that focus on systems for data management.

**EC6** Papers whose subjects is law.

The authors excluded papers that were related to pets (EC1), as well as studies that employed non-digital methods to collect PHD (EC2). Besides that, were also excluded studies concerning the assessment of medical information systems, namely studies that were focused only on the technical aspects of healthcare systems (EC3) and information systems interoperability (EC4). Likewise, out of the scope of this systematic mapping study was the investigation of practices in health management (EC5) and legal frameworks (EC6) as they are contextual to business-oriented studies.

In total, 246 papers were identified after the removal of duplicates. When the same paper appeared in more than one source, it was considered only once according to our search order. Thereafter, 167 studies were excluded based on the inclusion and exclusion criteria leaving for the final result 79 selected studies.

### 3.4. Data extraction strategy

The selected studies were exploited to collect the data that would provide the set of possible answers to the MQs. The publication source and channel of the papers selected respond to **MQ1**, while the publication year responds to **MQ2**.

**MQ3.** A research type can be classified in the following categories [53]:

- Evaluation research: Existing PHD approaches are implemented in practice and an evaluation of them is conducted.
• Solution proposal: A PHD solution is proposed. This solution may be a new PHD approach or a significant extension of an existing approach. The potential benefits and the applicability of the solution could be shown with an empirical study or a good argumentation.

• Opinion paper: These papers express the personal opinion of somebody whether a certain technique is valuable or not, or how things should been done.

• Review: Analysis of PHD existing literature.

• Other, e.g. experience papers, which express the personal experience of author(s), explaining the actions and practical aspects of the topic under discussion.

MQ4. The selected studies can be classified as a [54, 55]:

• Case study: An empirical inquiry that investigates a PHD approach within its real-life context.

• Survey: A method for collecting quantitative information concerning a PHD approach, e.g. a questionnaire.

• Experiment: An empirical method applied under controlled conditions, in order to evaluate a PHD approach.

• History-based evaluation: Studies evaluating PHD approaches in previously completed software projects.

• Theory: Non-empirical research approaches or theoretical evaluation of a PHD approach.
MQ5. An approach can be classified as [52, 56]:

- Process: A series of actions, or functions leading to a PHD result and performing operations on data.
- Method: A regular and systematic means of accomplishing PHD.
- Tool-based technique: A technique based on a software tool to accomplish PHD tasks.
- Model: A system representation that allows PHD to be investigated through a hierarchical structure.
- Framework: A real or conceptual structure intended to serve as a support or guide for PHD.
- Other, e.g. guidelines, data mining technique.

MQ6. A PHR-related topic can be classified as: PHD security, PHD privacy, PHD accessibility and PHD analysis among others that can be identified from the selected studies. To identify the main research topics of the papers, the authors relied on the analysis of the title and keywords. In cases of papers with two main topics, the authors classified the papers keeping both topics.

3.5. Synthesis method

The synthesis method was based on:

1. Counting the number of papers per publication channel and the number of papers found in each bibliographic source per year,
2. Counting the primary studies that are classified in each MQ’s response,
3. Presenting charts for the classification results which have been used in the analysis,
4. Presenting in the discussion a narrative summary with which to recount the principal findings.

4. Results

This section describes the results related to the systematic MQs presented in Table 1. Table 2 and Table 3 give an overview of the classification results.

4.1. MQ1. Which publication channels are the main targets for PHD research?

62% of the selected papers are published in scientific journals, while 28% are in conferences. Only few studies are published in workshops, symposia, books or other channels of publication. Table 4 presents publication sources which has published more than one paper included in this study. The four sources identified are all journals. The rest of papers not shown in Table 4 were published in different venues and journals.

4.2. MQ2. How has the frequency of studies related to PHD changed over time?

Fig. 2 presents the number of articles published per year from 1991 to 2017. As Fig. 2 shows, the number of publications was less than 3 papers per year in the period from 1991 to 2013, except the years 2009 and 2012, while from 2014 the number of publications rises steadily.
Table 2: Classification results. Part 1. Acronym: History-based evaluation (HbE)

<table>
<thead>
<tr>
<th>Ref.</th>
<th>P. Channel</th>
<th>Year</th>
<th>Research Type</th>
<th>Empirical Type</th>
<th>Contribution Type</th>
<th>Research topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>[61]</td>
<td>Journal</td>
<td>2003</td>
<td>Review</td>
<td>No</td>
<td>Method</td>
<td>Data use</td>
</tr>
<tr>
<td>[62]</td>
<td>Conference</td>
<td>2005</td>
<td>Solution proposal</td>
<td>No</td>
<td>Framework</td>
<td>Data representation</td>
</tr>
<tr>
<td>[63]</td>
<td>Conference</td>
<td>2006</td>
<td>Solution proposal</td>
<td>No</td>
<td>Model</td>
<td>Sensor data</td>
</tr>
<tr>
<td>[64]</td>
<td>Journal</td>
<td>2007</td>
<td>Evaluation research</td>
<td>Survey</td>
<td>Guidelines</td>
<td>Data privacy</td>
</tr>
<tr>
<td>[67]</td>
<td>Journal</td>
<td>2007</td>
<td>Evaluation research</td>
<td>HbE Tool</td>
<td></td>
<td>Data sharing</td>
</tr>
<tr>
<td>[68]</td>
<td>Conference</td>
<td>2008</td>
<td>Solution proposal</td>
<td>Experiment</td>
<td>Framework</td>
<td>Sensor data</td>
</tr>
<tr>
<td>[69]</td>
<td>Conference</td>
<td>2008</td>
<td>Solution proposal</td>
<td>No</td>
<td>Tool</td>
<td>Data visualization</td>
</tr>
<tr>
<td>[70]</td>
<td>Conference</td>
<td>2008</td>
<td>Solution proposal</td>
<td>Experiment</td>
<td>Tool</td>
<td>Data visualization</td>
</tr>
<tr>
<td>[71]</td>
<td>Journal</td>
<td>2008</td>
<td>Evaluation research</td>
<td>Survey</td>
<td>Method</td>
<td>Data privacy</td>
</tr>
<tr>
<td>[72]</td>
<td>Conference</td>
<td>2008</td>
<td>Solution proposal</td>
<td>No</td>
<td>Framework</td>
<td>Data privacy &amp; security</td>
</tr>
<tr>
<td>[73]</td>
<td>Conference</td>
<td>2009</td>
<td>Solution proposal</td>
<td>Experiment</td>
<td>Method</td>
<td>Data accessibility</td>
</tr>
<tr>
<td>[74]</td>
<td>Journal</td>
<td>2009</td>
<td>Review</td>
<td>No</td>
<td>Method</td>
<td>Data privacy &amp; security</td>
</tr>
<tr>
<td>[75]</td>
<td>Journal</td>
<td>2009</td>
<td>Evaluation research</td>
<td>Experiment</td>
<td>Tool</td>
<td>Data accessibility</td>
</tr>
<tr>
<td>[76]</td>
<td>Journal</td>
<td>2009</td>
<td>Evaluation research</td>
<td>Survey</td>
<td>Method</td>
<td>Health records</td>
</tr>
<tr>
<td>[77]</td>
<td>Journal</td>
<td>2009</td>
<td>Evaluation research</td>
<td>Survey</td>
<td>Model</td>
<td>Data sharing</td>
</tr>
<tr>
<td>[79]</td>
<td>Other</td>
<td>2009</td>
<td>Evaluation research</td>
<td>Experiment</td>
<td>Method</td>
<td>Data privacy</td>
</tr>
<tr>
<td>[80]</td>
<td>Journal</td>
<td>2010</td>
<td>Opinion paper</td>
<td>No</td>
<td>Guidelines</td>
<td>Data accessibility</td>
</tr>
<tr>
<td>[81]</td>
<td>Journal</td>
<td>2010</td>
<td>Solution proposal</td>
<td>No</td>
<td>Model</td>
<td>Data protection</td>
</tr>
<tr>
<td>[82]</td>
<td>Conference</td>
<td>2010</td>
<td>Solution proposal</td>
<td>Other</td>
<td>Framework</td>
<td>Sensor data</td>
</tr>
<tr>
<td>[83]</td>
<td>Conference</td>
<td>2010</td>
<td>Evaluation research</td>
<td>Experiment</td>
<td>Method</td>
<td>Data visualization</td>
</tr>
<tr>
<td>[84]</td>
<td>Symposium</td>
<td>2010</td>
<td>Solution proposal</td>
<td>Survey</td>
<td>Framework</td>
<td>Data sharing</td>
</tr>
<tr>
<td>[85]</td>
<td>Conference</td>
<td>2010</td>
<td>Solution proposal</td>
<td>No</td>
<td>Tool</td>
<td>Data management</td>
</tr>
<tr>
<td>[87]</td>
<td>Conference</td>
<td>2011</td>
<td>Solution proposal</td>
<td>Experiment</td>
<td>Tool</td>
<td>Data integration</td>
</tr>
<tr>
<td>[88]</td>
<td>Conference</td>
<td>2011</td>
<td>Solution proposal</td>
<td>Survey</td>
<td>Method</td>
<td>Data sharing</td>
</tr>
<tr>
<td>[89]</td>
<td>Journal</td>
<td>2011</td>
<td>Solution proposal</td>
<td>Experiment</td>
<td>Framework</td>
<td>Data portability</td>
</tr>
<tr>
<td>[90]</td>
<td>Conference</td>
<td>2011</td>
<td>Solution proposal</td>
<td>Experiment</td>
<td>Framework</td>
<td>Data design</td>
</tr>
<tr>
<td>[91]</td>
<td>Journal</td>
<td>2011</td>
<td>Review</td>
<td>No</td>
<td>Method</td>
<td>Data accessibility</td>
</tr>
<tr>
<td>[93]</td>
<td>Journal</td>
<td>2012</td>
<td>Solution proposal</td>
<td>No</td>
<td>Model</td>
<td>Data protection &amp; privacy</td>
</tr>
<tr>
<td>[94]</td>
<td>Journal</td>
<td>2012</td>
<td>Evaluation research</td>
<td>Survey</td>
<td>Method</td>
<td>Data sharing</td>
</tr>
<tr>
<td>[96]</td>
<td>Journal</td>
<td>2012</td>
<td>Evaluation research</td>
<td>Survey</td>
<td>Method</td>
<td>Health Informatics &amp; Education</td>
</tr>
</tbody>
</table>
Table 3: Classification results. Part 2. Acronym: History-based evaluation (HbE)

<table>
<thead>
<tr>
<th>Ref.</th>
<th>P. Channel</th>
<th>Year</th>
<th>Research Type</th>
<th>Empirical Type</th>
<th>Contribution Type</th>
<th>Research topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>[95]</td>
<td>Workshop</td>
<td>2015</td>
<td>Evaluation research</td>
<td>Experiment</td>
<td>Tool</td>
<td>Sensor data</td>
</tr>
<tr>
<td>[97]</td>
<td>Other</td>
<td>2015</td>
<td>Evaluation research</td>
<td>Experiment</td>
<td>Framework</td>
<td>Data privacy</td>
</tr>
<tr>
<td>[34]</td>
<td>Conference</td>
<td>2015</td>
<td>Review</td>
<td>No</td>
<td>Method</td>
<td>Data protection</td>
</tr>
<tr>
<td>[98]</td>
<td>Workshop</td>
<td>2015</td>
<td>Solution proposal</td>
<td>Experiment</td>
<td>Framework</td>
<td>Data sharing</td>
</tr>
<tr>
<td>[40]</td>
<td>Journal</td>
<td>2015</td>
<td>Review</td>
<td>No</td>
<td>Method</td>
<td>Data management</td>
</tr>
<tr>
<td>[99]</td>
<td>Journal</td>
<td>2015</td>
<td>Solution proposal</td>
<td>Case study</td>
<td>Tool</td>
<td>Data management</td>
</tr>
<tr>
<td>[100]</td>
<td>Journal</td>
<td>2016</td>
<td>Evaluation research</td>
<td>Survey</td>
<td>Method</td>
<td>Data sharing</td>
</tr>
<tr>
<td>[101]</td>
<td>Journal</td>
<td>2016</td>
<td>Evaluation research</td>
<td>Survey</td>
<td>Method</td>
<td>Data sharing</td>
</tr>
<tr>
<td>[102]</td>
<td>Journal</td>
<td>2016</td>
<td>Solution proposal</td>
<td>Experiment</td>
<td>Tool</td>
<td>Data privacy &amp; security</td>
</tr>
<tr>
<td>[103]</td>
<td>Conference</td>
<td>2016</td>
<td>Solution proposal</td>
<td>Survey</td>
<td>Framework</td>
<td>Data analysis</td>
</tr>
<tr>
<td>[104]</td>
<td>Workshop</td>
<td>2016</td>
<td>Review</td>
<td>No</td>
<td>Method</td>
<td>Data use</td>
</tr>
<tr>
<td>[106]</td>
<td>Journal</td>
<td>2016</td>
<td>Solution proposal</td>
<td>No</td>
<td>Method</td>
<td>Data privacy</td>
</tr>
<tr>
<td>[14]</td>
<td>Conference</td>
<td>2016</td>
<td>Solution proposal</td>
<td>No</td>
<td>Tool</td>
<td>Data management</td>
</tr>
<tr>
<td>[36]</td>
<td>Conference</td>
<td>2017</td>
<td>Opinion paper</td>
<td>No</td>
<td>Guidelines</td>
<td>Data privacy</td>
</tr>
<tr>
<td>[111]</td>
<td>Conference</td>
<td>2017</td>
<td>Evaluation research</td>
<td>HbE</td>
<td>Model</td>
<td>Data privacy &amp; security</td>
</tr>
<tr>
<td>[113]</td>
<td>Journal</td>
<td>2017</td>
<td>Evaluation research</td>
<td>Experiment</td>
<td>Tool</td>
<td>Health records</td>
</tr>
<tr>
<td>[37]</td>
<td>Conference</td>
<td>2017</td>
<td>Solution proposal</td>
<td>No</td>
<td>Method</td>
<td>Health records</td>
</tr>
<tr>
<td>[116]</td>
<td>Journal</td>
<td>2017</td>
<td>Review</td>
<td>No</td>
<td>Method</td>
<td>Health Informatics &amp; Education</td>
</tr>
<tr>
<td>[118]</td>
<td>Journal</td>
<td>2017</td>
<td>Evaluation research</td>
<td>Experiment</td>
<td>Framework</td>
<td>Data protection &amp; privacy</td>
</tr>
<tr>
<td>[119]</td>
<td>Journal</td>
<td>2017</td>
<td>Evaluation research</td>
<td>Case study</td>
<td>Model</td>
<td>Data management</td>
</tr>
</tbody>
</table>

Table 4: Publication sources which published more than one PHD selected paper

<table>
<thead>
<tr>
<th>Publication source</th>
<th>References</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal of Medical Internet Research</td>
<td>[66, 30, 100, 71]</td>
<td>4</td>
</tr>
<tr>
<td>European Journal of Risk Regulation</td>
<td>[112, 81]</td>
<td>2</td>
</tr>
<tr>
<td>International Journal of Medical Informatics</td>
<td>[107, 35]</td>
<td>2</td>
</tr>
<tr>
<td>Studies in Health Technology and Informatics</td>
<td>[108, 99]</td>
<td>2</td>
</tr>
</tbody>
</table>

15
4.3. MQ3. *What are the research types of PHD studies?*

Fig. 3 shows the research type of the selected papers. Around 41% of the selected papers were solution proposal studies, 33% of the selected papers were undertaken to evaluate PHD existing approaches, 11% were reporting the authors’ opinions regarding PHD and the remaining papers were classified as reviews. Fig. 3 shows also that 41% of the solution proposals were empirically validated and that the majority of the suggested solutions are methods.
4.4. MQ4. Are PHD studies empirically validated?

Fig. 3 shows if the selected studies were empirically validated, and presents the empirical types used in the validation of PHD approaches. 33% of the selected studies were not evaluated empirically. 24% of the selected papers were evaluated with experiments, 24% used surveys, 4% undertook case studies and 3% used history-based evaluation.

4.5. MQ5. What are the approaches that were reported in PHD research?

Fig. 3 presents the approaches used for PHD in the selected papers. The approaches most frequently reported are methods (29% of the selected papers) followed by frameworks (25%). Guidelines, tool-based techniques and models were also identified in the selected studies.

4.6. MQ6. What are the main topics in PHD literature?

Fig. 4 shows that the main research topic of the selected papers is data privacy, followed by data sharing and data security. Other research topics were also identified in PHD literature and are presented in Fig. 5.

5. Discussion

This section discusses the results and main findings of this paper. Recommendations for future research are also presented.

5.1. Principal findings

The main goal of this systematic mapping study is to examine the current research in PHD. The principal findings of this study are the following:
Figure 4: Research topics of the selected papers

Figure 5: Research topics in selected studies about Personal Health Data
5.1.1. Publication channels

Publication channels are informative about the process of development of PHD research area. The results obtained in this study show that PHD publications are mainly published in scientific journals (62%). Scientific publications in journals have been seen as an elaborate process in terms of review criteria, acceptance practices and manuscript-journal “fit”, so that the contribution of scientific publications in prestigious journals is not disputable [120, 121]. Furthermore, the visibility of publications in peer-reviewed journals is considered to be higher than other publication media [122]. Publication of peer reviewed studies in high-impact journals like JMIR, BMC and JMS, are indications of the latest growth of this research area across disciplines.

In opposition to journals, conferences have been characterized as innovation-laden, while are faster publication venues for dissemination of research results, promoting also knowledge exchange [123]. The discussion regarding the empirical types of studies bellow, will further strengthen this argument discussing the current types of research in this research area. The frequency of published studies over time, will be discussed extensively in the following subsection, but a primary observation is that PHD research is a relatively new scientific research area that started progressing over the last decade. An indicator of the recent development of this research area might be attributed to the lack of EHRs interoperability standards. According to WHO, 47% of the Members States worldwide have national EHRs, with the majority of them stating interoperability as one of the main challenges for the implementation of EHRs. The adoption rates of EHRs reported over 50% for
upper-middle and high-income countries, while for the lower-middle and low-income countries the rates are lower, 35% and 15% respectively [124, 125]. Nevertheless, the adoption of PHR found to dependent on growth in EHR adoption [17].

5.1.2. Publication trend

The present study classified publications from 1991 to 2017. The PHD research has attracted increasing attention since 2007. The lack of publications between 1991 and 2013 is reasonable as PHRs and EHRs are relatively recent developments in digital health. A primary definition of PHD was given by Markle Foundation’s Personal Health Working Group was in 2003 [126], while in 2005 in a working symposium, the College of Medical Informatics formalized a definition of PHD [8]. Later that year, the American Health Information Management Association (AHIMA) launched a Work Group to study the role of PHD in EHR [127]. Also, the AHIMA’s Foundation program “Better health information for all” started in 2009.

Personal health organizers like HealthSpace became available to all patients in UK in 2003, facilitating them to store personal health notes [128]. The adoption of EHRs in a national level in US, Australia and UK was initiated in 2004 [12]. While one of the first on-line communities for patients, the social platform PatientsLikeMe, was launched in 2006 [129], well-known PHRs such as Microsoft HealthVault [130] and Google Health [131] were released in 2007 and 2008 respectively. Nevertheless, Google’s decision to stop supporting Google Health application indicates that the popularity of the product was lower than anticipated. The use of EHRs is restricted to medical information exchange systems, making evident that the integration of
EHRs and PHRs is still in its infancy [125]. What considered valuable since PHRs’ conception, is the ability to connect personal records with EHR, as stand alone patients’ data are likely to remain outdated, acting as archives for patients’ input [8, 17].

Connected health emergence within the last decade along with technological advancements in healthcare aim to change health management of businesses and organizations [132]. According to Chouvarda et al., connected health “aims for the optimal access, sharing, analysis and use of health data via systematic application of healthcare information technology, in other words to “offer the correct information to the correct person at the correct time”, so that health actors (citizens, patients, clinicians, policy makers) make better decisions for health and care” [133]. Connected health includes e-Health, m-Health, telemedicine and telehealth solutions and refers also to the provision of health services remotely, with applications to the global aging population and to patients with chronic conditions [134]. The emergence of connected health model is bounded to a shift in global demographics; the share of aging population projected to increase surpassing young people [1]. The pressure of the aging population to global economy [135], comes as a result of an increasing population with chronic and degenerative diseases [1].

Reflecting upon these, we could argue that the growth of scientific publications in PHD area from 2014 onwards, portrays the demand for provision of better self-care services and adoption of agile solutions to facilitate the daily life of population. Besides that, the adoption of technological advancements such as smartphones brings patients closer than ever to personalized health care [136]. According to IDC, 344.3 million smartphones were sold
worldwide in the first quarter of 2017 [137], with the number of smartphone users forecasted to reach 2.1 billion in 2020 [138], while sensors connected to IoT are projected to range from 30 to 50 billion devices [139, 140]. In agreement with these trend is also the decision of Agfa Healthcare, one of the key-players in the Healthcare IT industry to ally with My Personal Health Record Express Inc. buying 27% of the company on 2016 [141].

5.1.3. Research types

The results show that the main concern of researchers in the PHD domain is to propose novel approaches or to extent existing techniques to deal with PHD. Moreover, evaluation research also is a practice used to assess previous research efforts by implementing a solution in practice, in order to provide practical explanations and to indicate advantages and disadvantages [46, 53, 142]. The results indicate that research in this area is still in its early stages, so it is essential to evaluate processes and solutions in order to evolve.

Around a third of the selected studies were not evaluation research, suggesting that the proposed approaches are not validated based on users experiences. These results suggest that future research should be focusing towards research approaches that will enable user validation in order to encapsulate users’ perceptions. The value of human-centered design cannot be neglected; especially in healthcare applications users involvement throughout the design process from low-fidelity prototypes to usability testing of the final products have been seen as a crucial factor for developing valuable medical systems in connected health [143]. The Agency for Healthcare Research and Quality, as well as Food and Drug Administration (FDA) have also discussed the importance of usability and human factors evaluation of medical devices and
applications during the design process [144, 145].

The following two categories, namely reviews and opinion papers are more theoretical approaches, as they do not rely on empirical validation of the proposed solutions [53, 142, 46]. The aim of these research approaches is to discuss or frame theories of PHD and to provide a theoretical framework.

5.1.4. Empirical studies

The findings for this MQ show that almost half of the selected studies are not validated empirically. This means, that the studies do not provide validation of results, while the support of the proposed approaches is not based on any proof of concept. More specifically, approximately 43% of the non-empirical studies suggest theoretical approaches revealing that there is a lack of user validation on the proposed solutions in real or laboratory settings. Our results also indicate that only 24% of the selected papers were evaluated with experiments. Despite this, almost all of the studies that were classified as experiments, were conducted in laboratory settings without the participation of users [28, 69, 75]. An explanation regarding this result is related to the context of the studies. The majority of them are experimental research on technical solutions such as data mining approaches [39], cloud computing [102], or data fusion and aggregation methods for pre-processing of PHD [42]. Note also that according to our classification scheme, an experiment is an empirical method applied under controlled conditions to evaluate a PHD approach. Therefore, there is no strict constraint for evaluating the proposed solution on users. Besides that, the same percentage with experiments was classified as surveys. Surveys are methods that capture the opinion of users indirectly using instruments such as questionnaires. The low percentage of
case studies also verifies the lack of user validation on the proposed solutions, as we only identify three studies: two solution proposals about sensor data and data management [78, 99], and one evaluation research which also concerns management of data [119]. The existence of these three case studies could be a good approximation of the implementation status of PHD in healthcare industry, as a case study validation requires real-world context [54, 55]. History-based evaluations also indicate that PHD research area is in an early development stage, as they point out the lack of investigation of the cases in real-world settings. Future research efforts should be focusing towards empirical validation of the proposed solutions.

5.1.5. Approaches

The contribution types of the selected papers can be seen in Fig. 3. The approaches most frequently reported are methods 29%, followed by frameworks 25%. Methods are used in research to report the steps for knowledge acquisition, while frameworks offer guidance for the development of solutions [52, 56]. The results of this mapping study show that the majority of the authors propose methods, however almost half of the cases are reviews and thus are not empirically validated [29, 34, 40, 61, 74, 96, 104, 116]. The approach of authors reveals that they only assess previous knowledge without proposing real implementation of PHD solutions. Only one study that discusses sensor technology, proposes a framework validated on data extracted from wearable sensors of a football team [78]. The contribution types of the selected papers point out a lack of user validation on the suggested solutions. The lack of user validation can be attributed to limited implementation of PHD solutions in medical industry. To our knowledge medical information
systems that allow integration of EHR and PHR are modest or restricted to pilot implementations [146]. Some of the PHD approaches reported in the selected papers are listed below.

Li et al. discuss access control of PHR by users in cloud-computing proposing a framework for secure access control and distribution of data in systems with multiple security domains [38]. Privacy of PHD in Big data technologies in healthcare is the subject of Lee et al. study [106]. The complexity of de-identification of personal information hinders the implementation of such technologies in medical sector. The root of problem is also spotted on the lack of common agreement among parties that distribute and handle data. Pickard’s study presents a global online survey about consumers’ willingness to share PHD, for the creation of datasets that can predict wellness and diseases trends [86]. The survey is focusing on the influence of the financial rewards towards information sharing within the context of Markets of Data (MoDAT). Participants’ willingness to share PHD has also been discussed from a different perspective by Chen et al. [101]. They concluded that participants’ willingness to share personal data depends on the potential data usage. The study points out that health research and the development of targeted interventions increased participants willingness to share their data, creating thus opportunities for both researchers and industry. PHD management is a research topic that has attracted the interest of researchers and practitioners from various disciplines, such as Engineers, Computer Scientists and Medical Professionals [14, 37, 113]. Mobile-based solutions for integration of PHD to healthcare systems [31], as well as a multi-platform synchronization framework useful for speeding up the imple-
5.1.6. Research topics

PHD is an interdisciplinary research area with contributions to various disciplines [33, 84, 115, 116]. The main research topics of the selected papers is data privacy (20%), followed by data sharing (12%) and data security (11%). What the present results indicate is that although the technology supporting PHRs is available [147], among the main concerns are privacy and security of PHD. The discussions of the selected studies are around these topics, focusing either on technical features of PHD systems like cryptography [33], cloud-computing and big data [89, 105], and security in mHealth [98] or on the legal and ethical aspects of it [57, 115]. These findings suggest that the research community approaches holistically privacy and security of PHD, identifying that both technical and human factors are essential for the adoption of sustainable solutions. Although this study is focusing on the technological aspects of PHD, the impact of the human factor on privacy and security solutions in healthcare cannot be neglected. What this study confirms is that technology alone is not enough to handle these challenges, suggesting the need for introduction of proactive legal and ethical frameworks.

Technology acceptance and adoption is a topic that has been discussed on previous studies in other research areas [148], but a general principle that applies also on PHD area is that the widespread usage of PHD requires the acceptance and adoption of both medical professionals and patients.

Many of the selected studies point out that investigations are also focusing on technical and users’ aspects concerning a future centralization approach of PHD (Data sharing 12%) [77, 83]. Data sharing is connected to privacy
and security, as without a secure framework it would be impossible to leverage patients’ PHD. Research types namely solution proposals and evaluation research, are categories with high number of publications concerning privacy and security. This indicates the need for proposition of novel solutions or significant extensions of existing approaches [149].

In contrast to data sharing, accessibility and data use were classified as less important influences on PHD area, probably because of limited real-life implementations in healthcare industry. However, access control of PHD is an issue of concern as counterweight of system preferences with security and privacy control found to be a hard task for patients [150]. Without security control over the data, patients are exposed to privacy threats [70]. The existing legal inconsistencies about personal data privacy and security will attempt to bridge the new General Data Protection Regulations (GDPR), which had been implemented in full on May 2018 [151]. The implementation of the new legislation in Europe is projected to have positive impact on data sharing in healthcare, while taking control of security and privacy issues [152]. On a worldwide scale according to WHO, 55% of the Member States have legal frameworks to protect the privacy of PHD, while 34% reported legal frameworks pertinent to sharing of PHD between health care professionals within the same country, and only 22% to sharing such data on an international level [125].

Although there has been a substantial increase in the number of policies and legislations for eHealth solutions the last decade, issues such as privacy, confidentiality and data security are among the main challenges that national healthcare systems around the world are facing as they evolve, making it
clear that the issue has not only legal applications but also ethical [124, 125]. The formulation of ethical frameworks from various organizations internationally addresses the need to trigger the ethos of the scientific community. The adoption of ethical principles on decision-making pertinent to health data management can safeguard public health and lead to future sustainable solutions [153].

An interesting finding is that although many studies have addressed patients with chronic conditions and disabilities as user groups that could benefit the most from PHD utilization [154, 155, 156], only three papers [103, 113, 117] of the selected studies have empirically evaluated approaches using data from patients with chronic conditions [30].

5.2. Implications

The findings of this systematic mapping study have implications for both researchers who are planning new studies in PHD and for practitioners who are working in connected health and would like to have an overview on the existent studies on PHD research area. The present study could be used as a benchmark for future research regarding PHD. Moreover, the PHD studies presented in this systematic mapping study might help researchers to identify research directions that they have not studied so far in order to improve the quality of research of this discipline.

For future research on PHD, greater presence in conferences should be considered in order to disseminate recent findings and to communicate scientific results, values and methods to researchers outside the discipline. More evaluation research should be carried out focusing on user validation in order to evaluate the proposed solutions based on real-world settings. Therefore, it
is advised that future solutions should include experiments and case studies. An interesting finding is that although the existing literature has identified patients with chronic conditions as one of the target groups that could benefit from PHD adoption, in very few cases studies researchers have evaluated the proposed solutions based on data from this group of patients. Thus, in future endeavors, PHD research should be directed towards validation of research findings involving patients.

5.3. Threads to validity

5.3.1. Construct validity

Construct threats to validity are related to the identification of primary studies [157, 158]. To limit construct threats while identifying and selecting as many as possible relevant studies, two researchers initially ran several iterations to test different strings of potential keywords. Three keywords were used in the final iteration, as the number of the returned results was broader using this research string. PHD is an interdisciplinary research field, so by using a broader search string, we have found and included publications from various research areas, limiting the possibility of excluding relevant studies. However, the list of publications might be incomplete, as alternative or additional terms might have altered the final selection of papers [159].

In this study the papers were identified consulting the following databases IEEE Digital Library, ACM Digital Library, Science Direct, SpringLink and Google scholar. The search was limited to the title, the abstract and the keywords of papers. To limit the risk of excluding relevant publications, the authors sough to use a search string limited to three keywords to include a broad selection of relevant literature in the first iteration. In the present
study, the authors did not consider databases containing publications from specific regions such as Latin America and the Caribbean (Literatura Latino Americana em Ciências da Saúde, LILACS), Africa (African Index Medicus, AIM) and Eastern Mediterranean Region (Index Medicus for the Eastern Mediterranean Region, IMEMR). However, the literature search has been conducted on four major global databases that are among the world’s most leading and comprehensive sources for scientific research hosting thousands of academic journals and conference papers. Also, the authors have searched for relevant literature using Google Scholar, which is a more generic bibliographic database that indexes full text or metadata of scientific literature across disciplines. Therefore, we believe that the majority of relevant studies in the discipline has been included in this systematic mapping study. Also, including publications that are not published in high JCR impact factor journals or CORE conferences gave a proportional representation of the published research in the discipline. Due to the fact that this study is a mapping study, the exclusion and inclusion criteria are only related to whether PHD is the topic of the publication or not.

To disregard threads related to accessibility of publications due to subscription limitations of universities’ libraries, both researchers agreed to mark the missing papers during the initial screening. Only two publications in total were not accessible by the authors. To overcome this barrier researchers requested the missing papers via email in ResearchGate platform. Previous research has also addressed, as a potential threat to validity, duplicated publications [160]. When duplications were detected, they were examined exhaustively to discover whether or not they were the same study. Although
parts of content were common to different papers, these papers are based on novel ideas or studies. The initial selection of articles performed by the first author, while the second author revised the results. In a few cases researchers had to argue about their decision to include or exclude a publication referring back to the established inclusion and exclusion criteria.

5.3.2. **Internal validity**

Internal validity concerns data extraction and analysis [158, 161]. The final selection and classification of articles was performed by the first two authors, while the third author revised the final version of this study and contributed addressing possible points of contention. The initial screening of papers conducted while authors were located in the same physical space, namely at IT University of Copenhagen, as the second other was conducting a research stay there during January 2018. Although authors where then located in different time zones, the initial person to person collaboration and alignment about research and data analysis methodology, contributed to the establishment of effective communication guidelines, alleviating therefore tiredness or misunderstandings due to long-distance collaboration [162]. To limit the authors’ influence on the classification process, the development of the classification scheme relied on widely accepted and well-established guidelines [46]. In addition, what considered to be beneficial for the validity of the search strategy and results was the previous knowledge and skills of the second co-author on conducting systematic mapping studies [55, 56, 163, 164]. Internal validity refers also to the analysis of data. The analysis of results in this study relied only on descriptive statistics, therefore the threads to validity are minimal.
5.3.3. Conclusion validity

Conclusion validity relates to imprecise conclusions that might be drawn in case of identification of unreasonable relationships. In mapping studies this thread concerns inaccurate data extraction or exclusion of relevant studies during the selection phase [161]. The control over conclusion validity has also implications to reliability of the study, so that if a systematic mapping study is performed by other researchers, they will arrive at the same conclusions [158, 159]. To handle this thread, the selection and extraction process of the data performed according to the established scheme as described in Section 3. Furthermore, the discussion of results relied exclusively on data visualization, such as bubble plots and histograms, that were generated for all of the MQs. In our viewpoint, slight differences based on publication selection bias and misclassification would have minor impact on the main conclusions drawn from the 79 publications identified in our mapping study.

5.3.4. External validity

This thread refers to the generalization of this mapping study [165, 166]. The results of this study are considered with regard to the PHD domain, while the validity of conclusions is applicable only on the PHD context. The external validity threads are thus not applicable. Moreover, since no time range was specified to limit the search of the published papers, the representativeness of the selected studies was not affected. Therefore, this threat is not applicable in this context. The search string and the classification scheme presented in this paper can be perceived as a baseline for future research endeavors and practitioners can search for and categorize additional papers accordingly.
6. Conclusions and future work

This paper has presented a systematic mapping study that summarizes the existing research in PHD. Of 246 studies, 167 papers were identified between 1991 and 2017, 79 of which were selected and classified according to six criteria: publication source, publication year, research types, empirical types, contribution types and research topic. The findings showed that an increasing amount of attention has been paid to PHD since 2014. The two main research types found were solution proposals and evaluation research. The majority of the selected papers were empirically evaluated. Many papers proposed solutions for future research endeavors, and the main contribution types were methods and frameworks. Data privacy is the most frequently addressed topic in PHD literature, followed by data sharing. A few authors conducted research on PHD data processing or analysis. For future work, we intend to perform an SLR to identify and evaluate evidence research in PHD by taking into account the results revealed in this systematic mapping study to examine published literature in greater depth.

Author contributions

All authors contributed to the creation of the manuscript. MK and SO: design, conception, acquisition and interpretation of data, drafting of the manuscript, revision. MI contributed to the study by her critical revision. All authors read and approved the final manuscript. The order of authors listed in the manuscript has been approved by all of the authors.
Summary points

What was already known on the topic:

- There is an increasing interest about digital health hence the interest in PHD.
- No systematic mapping study has been conducted to gain insight into PHD.

What this study added to our knowledge:

- The publication trend of PHD research shows that there is an increasing interest in PHD since 2014.
- The main publication channel of PHD articles is journals.
- Solution proposals and evaluation research are the main research types of PHD literature.
- The majority of the selected papers were empirically evaluated.
- Methods and frameworks are the main approaches used in PHD literature.
- Data privacy is the most frequently addressed topic in PHD literature, followed by data sharing.

Conflict of interest

The authors confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Compliance with Ethical Standards

This article does not contain any studies with human participants or animals.

Acknowledgements

None.

References


[28] P. Fens, M. Funk, Personal health data: visualization modalities and their perceived values, in: Proceedings of the 22nd international confer-


[34] M. Mammadova, The Problems of Information Security of Electronic Personal Health Data, in: 7th International Conference on Information


[66] J. H. Frost, M. P. Massagli, Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another’s data, Journal of Medical Internet Research 10 (2008) e15.


[80] J. Gladwin, Patients welcome access to online health records: Jacqui Gladwin argues that giving people the right to see their personal medical data benefits them and the healthcare professionals involved, Primary Health Care 22 (2012) 10–10.


29th International Conference on Advanced Information Networking and Applications Workshops (WAINA), IEEE, 2015, pp. 411–416.


and Personal Health Information Management, CoRR abs/1606.09589 (2016).


[166] D. M. Rousseau, J. Manning, D. Denyer, 11 evidence in management and organizational science: Assembling the field’s full weight of scien-