

Remote Collaborative Knowledge Discovery for Better Understanding of Self-tracking Data

Lauri Tuovinen
Dublin City University
Dublin, Ireland

University of Oulu
Oulu, Finland
lauri.tuovinen@oulu.fi

Alan F. Smeaton
Dublin City University
Dublin, Ireland
alan.smeaton@dcu.ie

Abstract—Wearable self-tracking devices are an increasingly popular way for people to collect information relevant to their own health and well-being, but maximising the benefits derived from such information is hindered by the complexity of analysing it. To gain deeper insights into their own information generated by such products, a user with no data analysis expertise could collaborate with someone who does have the required knowledge and skills. To achieve such a successful collaboration, several tasks need to be completed: finding a collaborator, negotiating the terms of the collaboration, obtaining the necessary resources, analysing the data and evaluating the results of the analysis. To support the execution of these tasks, we have developed and deployed an online software platform that enables data collectors and owners to find experts and collaborate with them so they can extract additional knowledge from the self-tracking data. The functionality and user interface of the platform are demonstrated by presenting an application scenario where a data owner shares their sleep data with an expert who applies periodicity analysis to discover cyclical patterns from the data.

I. INTRODUCTION

The popularity of personal wellness products that collect and analyse data on the wearers' physiological state is increasing. These products enable a user to measure and log phenomena that are relevant to their health and well-being, such as physical activity and sleep. Even relatively inexpensive wearable products generate a diverse range of outputs and present them in an engaging and user-friendly way, so it is not surprising that such self-tracking products are gaining popularity, even though there are issues concerning their accuracy and the ability of the users to control their own data [1].

Personal data is widely recognised as a valuable resource that can be mined for useful knowledge using computational models. As a result of corporations exploiting personal data without adequate regard for the privacy of the individuals concerned, mining of personal data is frequently viewed in a negative light. However, this data could also be valuable to the individuals themselves as a resource and not merely as an implicit currency used to pay for access to services such as social media platforms.

Extracting the knowledge hidden in data patterns requires expertise that most people do not have and cannot be reasonably expected to acquire, yet the nature of self-tracking data is

such that collecting and analysing it is potentially beneficial to anyone regardless of background. To get around this problem, a person with data to analyse but no data analysis expertise could share it with someone who does have the required expertise and is willing to collaborate. To facilitate such collaboration between data owners and experts, we have developed an online software platform that enables its users to exchange datasets and analysis results and specify constraints that determine how much of their data is shared with each collaborator.

The software platform is a cloud-based client-server application with an underlying ontology that serves as a repository of domain knowledge and a shared data structure in which information about collaborators, collaborations and collaboration artefacts (e.g. datasets and data visualisations) is stored. The software and the underlying ontology have been previously described in [2], [3]; these papers view the software platform from a technical perspective, providing details on its architecture and implementation. In this paper we adopt an application-oriented perspective instead, examining the process of collaborative knowledge discovery and the role of the platform in it in the context of a concrete example of collaboration involving sleep data.

To provide evidence of the validity of the platform as a collaboration tool that can help anyone make sense of their self-tracking data, we discuss the principal tasks undertaken by collaborators and the support provided by the platform for the execution of these tasks. The main scientific contributions of the paper are thus:

- A description of the process steps in remote collaborative personal analytics, the requirements for supporting the steps and the functionality of the software platform in relation to these requirements;
- An application scenario demonstrating the process in action through a simulated collaboration where the software platform is used to share and study a real-world time series dataset of sleep data captured using a wearable self-tracking device.

Although simulated, the scenario proves that the platform can be used to carry out meaningful collaborations between data owners and knowledge discovery experts, and thus that the chosen approach is feasible. The results achieved so far

provide a foundation for future studies addressing the numerous questions that remain open in this new area of research.

The remainder of the paper is organised as follows: Section II provides background information and an overview of related work. Section III presents a process model for collaborative knowledge discovery from personal data and discusses its implications for software development. Section IV demonstrates how the software platform can be used to collaborate on periodicity analysis of sleep data, starting with the creation of a new collaboration and finishing with a visualisation of the analysis results. Section V presents a critical discussion of the results, and Section VI concludes the paper.

II. BACKGROUND

There is a wide variety of self-tracking products available in the market today, ranging from simple pedometers to high-end smartwatches. As a rule, more expensive devices have more on-board sensors and more advanced software for processing sensor readings, and consequently, more diverse outputs. However, even fairly low-end products are often capable of estimating variables such as distance, activity intensity, energy expenditure, sleep time and quality. As a product is used over a period of time, the recorded values build up into a time series from which various trends and other useful patterns may be discovered.

Typically, the data recorded by a self-tracking device is uploaded to the device manufacturer's cloud server, where the user can access it using a web browser or mobile application. Some types of time series analysis may be provided by the application, but if the user wishes to see analyses that are not among the features of the application, the only option is to export the data and study it using a more general-purpose data mining tool. Most users are not likely to be able to do this unless they can enlist someone skilled in the application of such tools to collaborate with them, assuming they themselves have not got those skills.

One type of analysis that might be of interest to many users of self-tracking products but is usually not provided by the products is periodicity analysis [4]. This type of analysis is highly relevant to self-tracking, because it provides a way to explore the cyclical patterns that occur in many human activities and physiological functions such as sleep. From sleep data sampled once per day, for an individual working a regular seven-day week, one would expect to find a periodicity corresponding to a seven-day cycle, but analysis may reveal other cycles for which there is no immediately obvious explanation. Discovering such patterns in one's sleeping habits, for example, may eventually lead to a better understanding of these patterns, which in turn may ultimately lead to actionable knowledge that can be used to enhance one's well-being through better sleep.

Several self-tracking products enable the user to download their data in a file, which may be in an open format such as CSV or a proprietary but widely supported one such as Excel. This provides a reasonably convenient means of sharing data with a collaborator, but when a collaboration takes place over a distance, additional tools are required to facilitate it. General-purpose applications such as email and instant messaging clients can be used to communicate and exchange

files remotely, which fulfils minimum requirements, but to collaborate more effectively, an online platform specifically designed to support this type of collaboration is required.

Numerous collaborative platforms for data analysis have been developed, although many of these have been designed for application only in a relatively narrowly-defined domain (e.g. [5], [6], [7], [8]). The most generic platforms can be broadly divided into two major categories: some support collaborative development of knowledge discovery workflows (e.g. [9], [10]), while others adopt a more ad-hoc, exploratory approach through collaborative notebooks (e.g. [11], [12]). In both types of systems, the participants in a collaboration can synchronously edit the collaborative artefacts by remotely connecting to the collaboration platform from their current location.

The central problem with the existing collaboration platforms in the context of personal data analytics is that they have not been designed with this special case in mind. In collaborative personal analytics, the collaboration process is driven by an individual who has data but not necessarily any technical expertise; this introduces additional complications, as does the potentially sensitive nature of the data. The purpose of our collaboration platform is to address the challenges characteristic of personal analytics. In the next section we take a closer look at what these challenges are and how we approached them.

III. SUPPORTING COLLABORATIVE KNOWLEDGE DISCOVERY

Our model for remote collaborative knowledge discovery views the process from the data owner's point of view. In this model, the process consists of five principal tasks:

- 1) **Finding collaborators:** identifying people with the right kind of expertise and inviting them to join your collaboration;
- 2) **Negotiating terms:** establishing the objectives of the collaboration and the contributions and expectations of each collaborator;
- 3) **Obtaining resources:** securing the datasets, tools and other non-human resources required to achieve the objectives;
- 4) **Performing analyses:** carrying out the data processing operations required to extract the desired knowledge from the data;
- 5) **Evaluating results:** interpreting the results of the analysis and assessing their significance.

Supporting the finding of collaborators is important because without such support, collaboration would only be accessible to individuals who already know someone with the required knowledge and skills. Our collaboration platform enables users to connect and communicate with others, regardless of location. In our platform, information about users' expertise is stored in an ontology and the software can be used to search this information and send collaboration invitations to candidates who appear suitable. Each user of the platform has a discoverability attribute that determines whether the user can be found by other users searching for collaborators; an undiscoverable user will not appear in any searches but can perform searches and send invitations to discoverable users.

Negotiating the terms of the collaboration is arguably the most important part of the process from the data owner’s perspective, since this is where the data owner is likely to have the biggest impact on the outcome of the collaboration. Fundamentally, the determination of the collaboration objectives is driven by the interests of the data owner. To support negotiation, the collaboration platform must minimally supply the collaborators with a communication channel, but ideally, it should help the collaborators arrive at a mutually satisfactory set of terms and conditions. In particular, the data owner should be supported in understanding the implications of the agreement, to compensate for the imbalance of power that may result from the collaborators having different levels of expertise.

To frame this point in more concrete terms, among the things to be negotiated are the boundaries of the data owner’s privacy. Assuming that the data owner is not willing to disclose all of their data with the expert collaborators (whom the data owner may not know very well), a trade-off needs to be found where the data owner’s privacy preferences are balanced against the data requirements of the analysis to be carried out. Since the data owner cannot be expected to be familiar with the requirements and privacy implications of data mining techniques, the collaboration platform should help them understand these.

In our platform, the negotiation of data sharing begins with an expert specifying a data request for a given dataset, to which the owner of the dataset can respond by specifying a set of constraints that limit the amount of data disclosed to the expert. In the current implementation of the software, the constraints simply block access to some elements of the requested dataset, but as discussed in [2], [3], the ontology enables representation of constraints where the data undergoes a transformation that reduces its sensitivity while retaining an acceptable level of utility for the desired analysis. Ultimately, the domain knowledge stored in the ontology should enable the software to suggest possible resolutions to conflicts between data requirements and privacy preferences.

Like collaborators, the non-human resources used in collaborative knowledge discovery may be geographically distributed, and therefore the collaboration platform should ideally enable users to discover and integrate such resources. Examples of potentially useful distributed resources include open datasets in online repositories and data mining algorithms implemented as web services. In our platform, resource discovery is currently not implemented, but locally available data resources can be imported in widely used portable formats such as CSV and JSON. An API is provided for expanding the implementation with new types of data sources. The number of choices to be made by the user when importing a dataset can vary depending on the user’s level of expertise; for example, an expert user may opt to set the data types of variables manually, whereas a novice user will probably prefer to let the software handle this.

For the data analysis task, the existing platforms mentioned in the previous section provide tools such as collaborative editors for analytics notebooks or knowledge discovery workflows. These are good for facilitating collaboration among experts, but they bring little or no added value to a collaboration between an expert and a non-expert, since the non-expert is

unlikely to be able to make any meaningful contributions using such tools. Conceiving a tool that would enable a non-expert to participate in this part of the process is a major challenge, and in our platform this is left out of scope in the current implementation, although the specification of data analysis operations is taken into account in the design of the underlying ontology. Instead of using the platform to process data, the expert can export the data in a portable format and process it using external tools deemed suitable for the job at hand. The results of the analysis can be imported into the platform as a derivative dataset, causing them to be automatically shared with the owner of the original input dataset.

In the evaluation of results the data owner will again be heavily involved, since the success of the collaboration is ultimately determined by whether the data owner is able to understand and apply the discovered knowledge. To help the expert explain results to the data owner, the collaboration platform should enable the expert to create various types of visualisations that the expert and data owner can explore together. In our platform, the expert can currently choose from a range of two-dimensional visualisations, which can be attached to datasets and viewed by other participants of the same collaboration who have access to the visualised dataset.

In addition to the functionality described above, our platform provides a chat function that can be used by the participants of a collaboration to communicate with one another. The chat is the only part of the software that does not use the underlying ontology: all other collaboration activities are handled internally as creation and deletion of axioms in the ontology. Each user has access only to those elements of the ontology that concern the user – for example, only those collaborations that the user is participating in or has been invited to – which reduces the amount of ontology data that needs to be transferred and protects the privacy of the users. Another privacy-preserving feature is that datasets are maintained locally by each data owner and uploaded to the server only temporarily when shared with a collaborator. More information on the architecture and implementation of the platform can be found in [3].

The steps of the process outlined above, the requirements of each step and the support provided by our platform for each step are summarised in Fig. 1. It is worth noting that in the real world the process does not always proceed in this linear, waterfall-style fashion, and the model is not intended to imply that it should. The main purpose of the model is to help us understand what features a collaboration platform should have in order to comprehensively support the process of collaborative knowledge discovery from personal data. In the next section we examine a complete execution of this process using the current implementation of our platform.

IV. APPLICATION SCENARIO

The scenario described below represents a simulated collaboration between a data owner and a time series analysis expert, aiming to discover periodicities in the data owner’s sleep data. The dataset used is a real-world data captured by one of the authors using a wearable self-tracking device. The simulated collaboration consists of the following steps:

- 1) Creating a new collaboration;

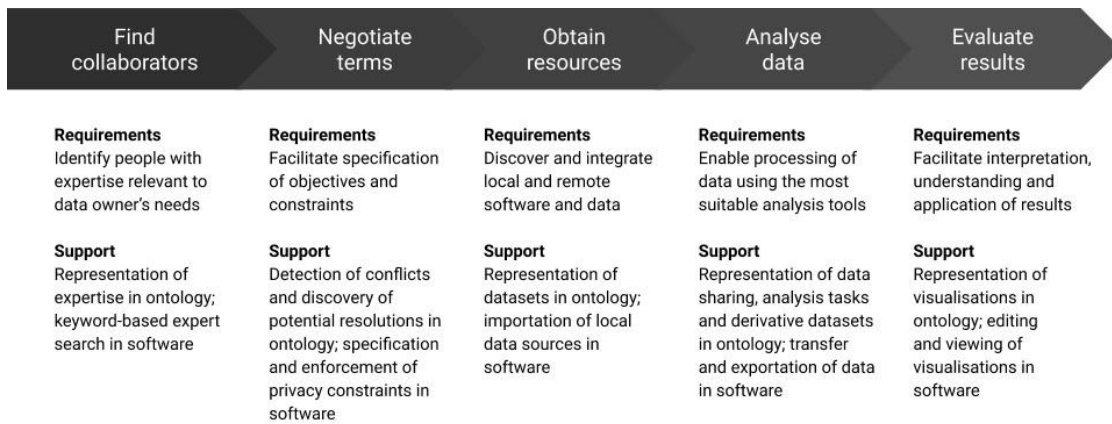


Fig. 1. Tasks to be carried out in collaborative personal analytics, the support required by each task and the relevant features of the proposed software platform

- 2) Creating a dataset and adding it to the collaboration;
- 3) Finding an expert and inviting them to the collaboration;
- 4) Sharing the dataset with the expert;
- 5) Analysing the data and adding the results to the collaboration as a dataset;
- 6) Creating a visualisation of the results and sharing it with the data owner.

When a user logs in to the software platform, the initial view displayed to the user is the main window, shown in Fig. 2, which presents the user with four tabs: Collaborations, Data management, My profile and Invitations. By clicking on the New collaboration button on the Collaborations tab (Fig. 2a), the user can initiate a new collaboration; the software will prompt the user for a name and a description for the new collaboration and place the name in the list on the left-hand side of the interface. The names of collaborations started by other users that the user is participating in will also be shown here. When the user selects a collaboration from the list, its description and other details will be displayed in the text pane on the right-hand side.

On the Data management tab (Fig. 2b) the user can create a new dataset by clicking on the New dataset button; as with collaborations, the user will first be prompted to enter a name and a description for the dataset. After this, the user will be given the option to import data into the new dataset using the import wizard shown in Fig. 3. The wizard first asks the user to specify whether the data to be imported comes from a wearable self-tracking device or some other source; further options will be shown depending on the user's input. If the data source is a wearable device (which is the case here), the user can simply select the type of device from a drop-down list and the software will select the remaining options on the user's behalf. The user can then proceed to click on the Open button to select the data file, and finally the Import button to read the file and import the data. Here, the user is importing data from a JSON file containing data collected using the Oura Ring activity and sleep tracker [13].

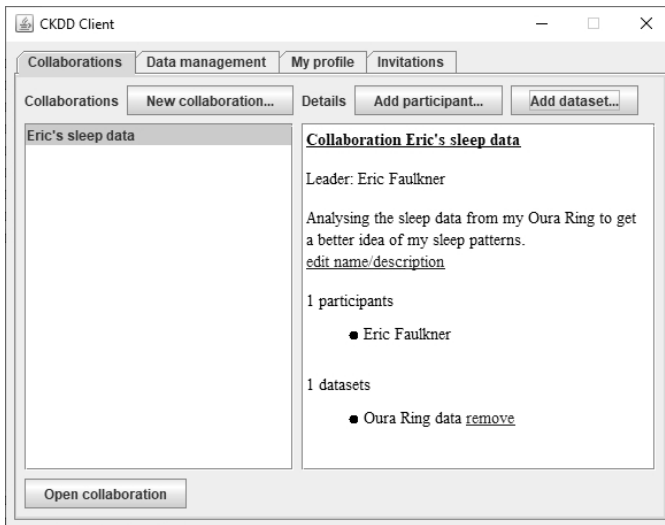
After a dataset has been created, it can be added to the collaboration by clicking on the Add dataset button on the Collaborations tab. The name of the dataset will then be displayed in the collaboration details pane when the collaboration

is selected. After this, the user can proceed to invite an expert to join the collaboration by clicking on the Add participant button, which will launch the dialog shown in Fig. 4. Here the user can search for experts by entering a set of keywords and clicking on the Search button. The experts found will be shown in the list below the query box, ranked according to how closely their expertise matches the query. The closeness of the match is displayed next to the expert's name as a number between 0 and 1. When the user selects an expert from the list, details on the expert will be shown in the text pane on the right-hand side. When a suitable expert has been identified, the user can type a message for the expert and click on the Send invitation button to invite the expert to the collaboration.

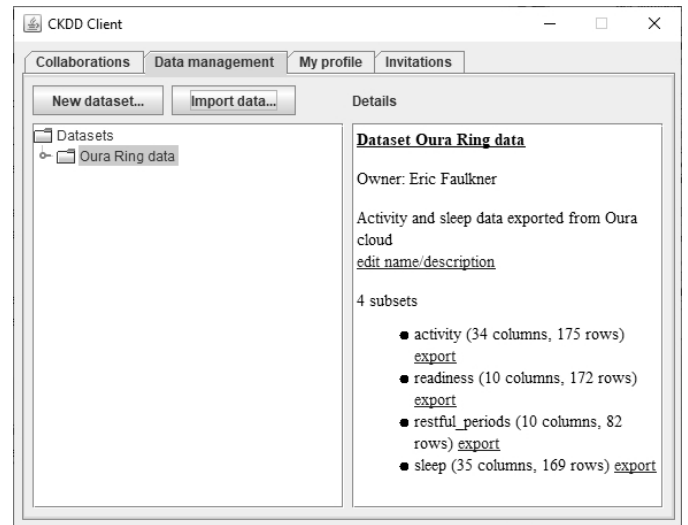
Sent and received collaboration invitations can be viewed on the Invitations tab (Fig. 2d) of the main window. When the user selects an invitation from one of the tables, its details will be displayed in the text pane below. On the right-hand side of the details pane, different buttons will be displayed depending on the type and status of the selected invitation. In the case of a received invitation to which the recipient has not yet responded (i.e. the status of the invitation is "pending"), the user can accept or reject the invitation by clicking on the Accept or Reject button. When the recipient has accepted the invitation, its status will change to "accepted" and the sender can click on the Add to collaboration button to close the invitation and add the recipient as a participant. The new participant's name will then be displayed in the collaboration details pane on the Collaborations tab when the collaboration is selected.

When the user selects a collaboration on the Collaborations tab and clicks on the Open collaboration button, the software will launch the collaboration window, shown in Fig. 5. Clicking on the Chat button opens the chat window (Fig. 6), where the users participating in the collaboration can discuss any matters that arise during the collaboration. The datasets used in the collaboration are shown in the list in the top-left corner of the collaboration window; to gain access to a dataset owned by another user, a participant can select a dataset from the list and click on the Request access button, which will create a new data request for the data owner and place it in the table in the bottom-left corner of the window.

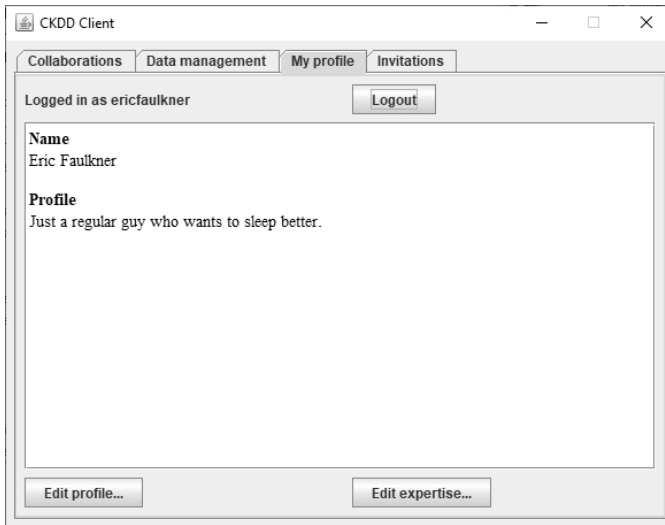
When the recipient of a data request selects a request from the table and clicks on the Review request button, detailed



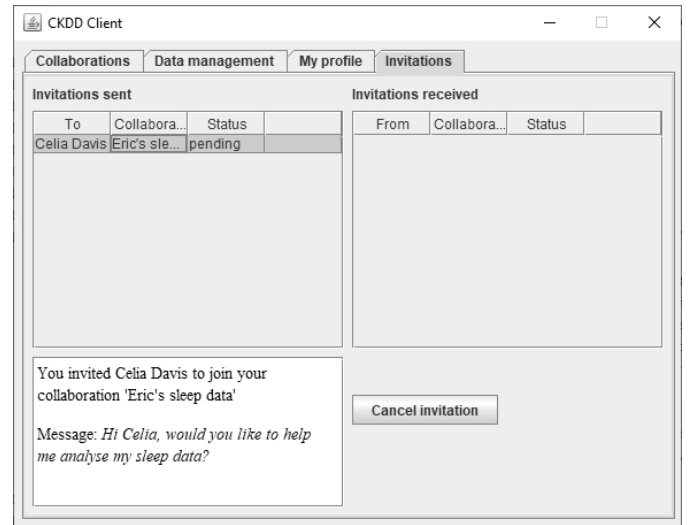
(a) Collaborations tab



(b) Data management tab



(c) Profile tab



(d) Invitations tab

Fig. 2. The four tabs of the main window of the collaborative knowledge discovery software platform

information about the request will be displayed in the dialog shown in Fig. 7a. The recipient can attach privacy constraints to the request by clicking on the Edit constraints button, which will launch the dialog shown in Fig. 7b. Here the recipient can select or de-select entire subsets or individual variables depending on how much data they wish to share with the sender of the request; the information in the text pane in the review dialog will then be updated to reflect the selections. When the recipient is satisfied with the constraints, they can click on the Grant request button, which will close the review dialog, upload the data to the server and notify the sender that the request has been granted.

When the sender of a data request selects the request and clicks on the Review request button, a different review dialog will be launched, shown in Fig. 7c. If the request has been granted and the data uploaded to the server, the Download data button will be enabled and the sender can click on it to receive

a copy of the dataset (or those parts of it that the data owner has chosen to share). Once the data has been downloaded, it can be exported by clicking on hyperlinks that will be shown in the details pane of the Data management tab of the main window when the shared dataset is selected. The exported data will be saved in a CSV-formatted text file, which can then be imported for further processing into any of the vast range of data analysis tools that support this format. In this scenario, the Spyder environment [14] and the SciPy libraries [15] were used to compute periodograms, which can be used to detect periodicities in self-tracking data [4].

When data analysis has been completed, the results can be added to the collaboration by clicking on the Create derivative button in the collaboration window. This will trigger a process that is otherwise similar to the creation of a new dataset, but additionally, in the underlying ontology the resulting dataset will be asserted as a derivative of the input dataset. From this

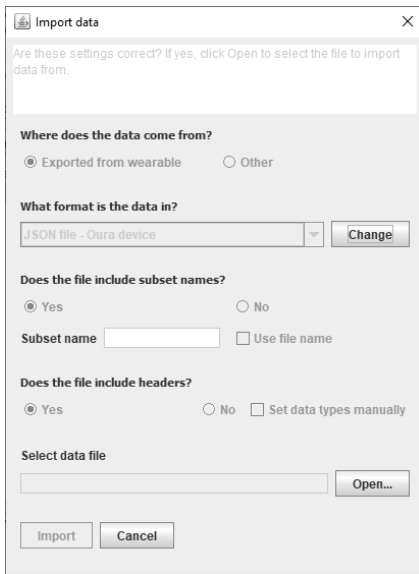


Fig. 3. Importing data into a dataset using the data import wizard

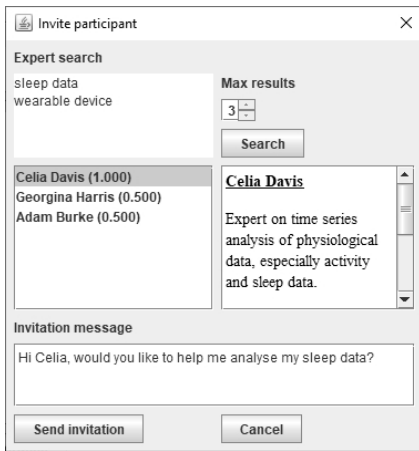


Fig. 4. Searching for experts and sending an invitation to collaborate

the software will infer that the owner of the input dataset should also have access to the result dataset, so the results dataset will be shared automatically without an explicit data request. The data owner will be notified when the results have been uploaded to the server and prompted to download them.

When the result data has been created and shared, the expert who performed the analysis can create a visualisation of it by selecting the dataset and clicking on the Create visualisation button in the collaboration window. This will first open a small dialog where the expert can select the type of visualisation to be created from a drop-down list. When a selection has been made, the software launches a visualisation editor, where the expert can select the data to be visualised and specify the parameters of the visualisation. Fig. 8 shows the visualisation editor, where a line chart visualisation of a periodogram is being created, showing a peak in spectral density at a frequency corresponding to a weekly cycle. When the visualisation has been created, anyone participating in the collaboration who has access to the underlying data can view

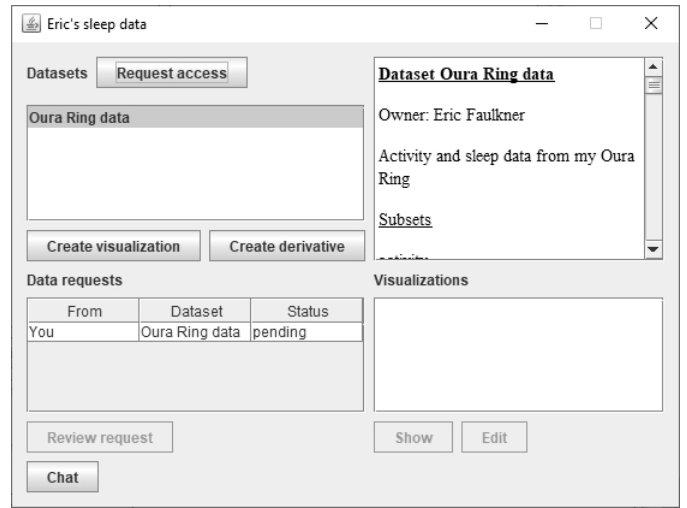


Fig. 5. The collaboration window, where the expert has just requested access to a dataset

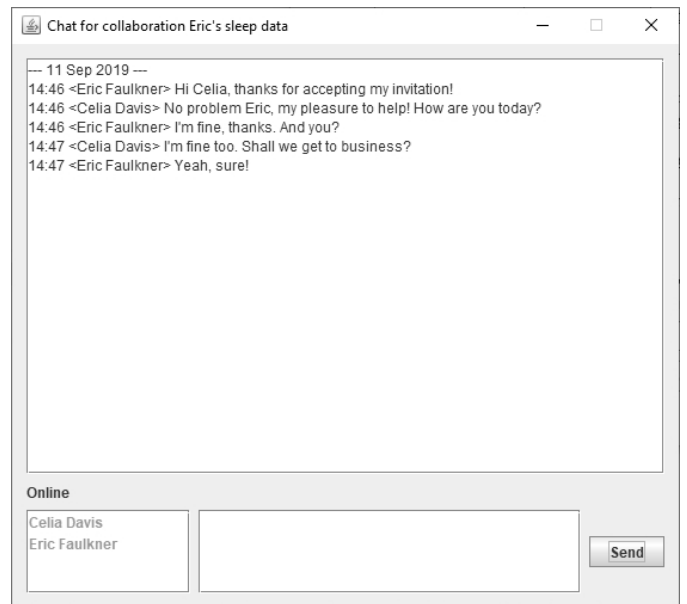
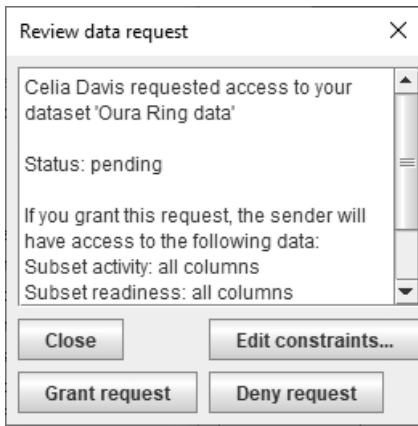


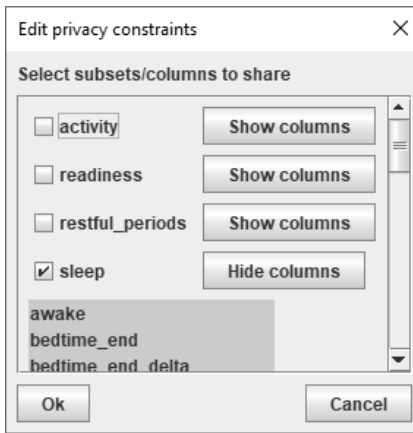
Fig. 6. Chatting with collaborators

it by selecting it from the list in the bottom-right corner of the collaboration window and clicking on the Show button.

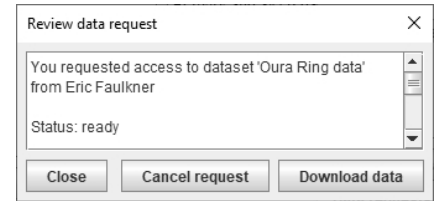
Fig. 9 presents a summary of the application scenario in the form of sequence diagrams. In Fig. 9a, the data owner logs in to the server, starts a new collaboration and invites an expert to join it; the initial sequence of sending login information to the server and requesting a copy of the ontology is repeated every time a user launches the client and connects to the server, but is only shown here once. In Fig. 9b, the data owner creates a dataset and adds it to the collaboration; the expert then requests access to the dataset, which the data owner grants with some constraints. Finally, in Fig. 9c, the expert exports the data for analysis, imports the results as a derivative dataset and creates a visualisation of this dataset; the data owner downloads the results and views the visualisation.



(a) Data owner reviewing the request



(b) Setting privacy constraints



(c) Expert reviewing the request

Fig. 7. Completing a data request

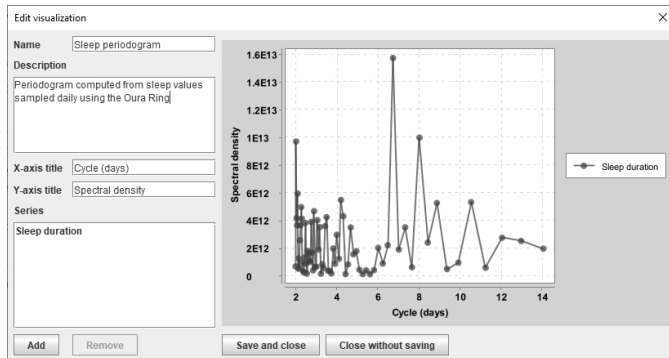


Fig. 8. Creating a visualisation

V. DISCUSSION

There are two major obstacles that are currently making it difficult for individuals to benefit fully from the useful knowledge that lies hidden in their data. The first one is that the data is mostly controlled by organisations that have a vested interest, typically a financial one, in preserving the status quo. However, the justification for the power of such organisations to control personal data is increasingly being questioned, and initiatives such as MyData Global [16] are aiming to tip the balance of power in favour of the individuals. Legislation affirming individuals' rights of determination concerning the processing of their personal data is also being enacted and enforced, the General Data Protection Regulation [17] of the European Union being the most prominent example.

The second obstacle is that extracting the useful knowledge from the data involves the application of data analysis techniques that require special skills to use. In theory, such skills can be acquired by anyone with access to the Internet, since there are free tools and learning materials available online, but in practice, in many cases the only realistic option is to collaborate with someone who has the required skills already. The potential impact of a platform capable of successfully facilitating such collaborations is thus considerable.

In fact, given the importance of physical activity and sleep to human health, the potential impact of any technology that helps people gain actionable insights into their self-tracking data is substantial. The application scenario described above shows that even with data coming from a single source, collaboration enables the data owner to go beyond the analyses provided out of the box by the self-tracking product used, but even greater potential comes with the possibility of combining data from multiple sources. Since the data is controlled by organisations rather than the data owners themselves, this is currently not as convenient as it ideally would be, but as long as the data can be downloaded from the data controllers' servers in a portable format, it can be imported into our software, which can then be used to manage all datasets regardless of how they were originally captured.

Since the collaboration in the application scenario is simulated, the strength of the evidence it provides in support of the software platform is limited. However, the scenario proves that all of the steps of the collaboration process as outlined in Section III can be carried out using the platform. By using the platform to chat and to exchange input datasets and analysis results, the data owner and the knowledge discovery expert can engage in an ongoing dialogue where they progressively improve their joint understanding of the data owner's data through a series of iterations. The end of the scenario as presented here therefore does not necessarily mark the end of the collaboration but rather the beginning of another iteration.

Remote collaborative knowledge discovery of the type discussed here, where the collaboration is driven by a non-expert individual hoping to extract useful knowledge from their own personal data, is still a relatively new concept, and there are many research problems in this area that require further attention. Perhaps the most difficult one is how to facilitate effective collaborative data analysis between experts and non-experts, where a clear vision has yet to emerge concerning how meaningful participation in data analysis activities without relevant expertise could work in theory, let alone in practice. Arguably, for the non-experts the reason for collaborating is specifically to avoid doing data analysis activities, but at the very least the collaboration platform should enable multiple

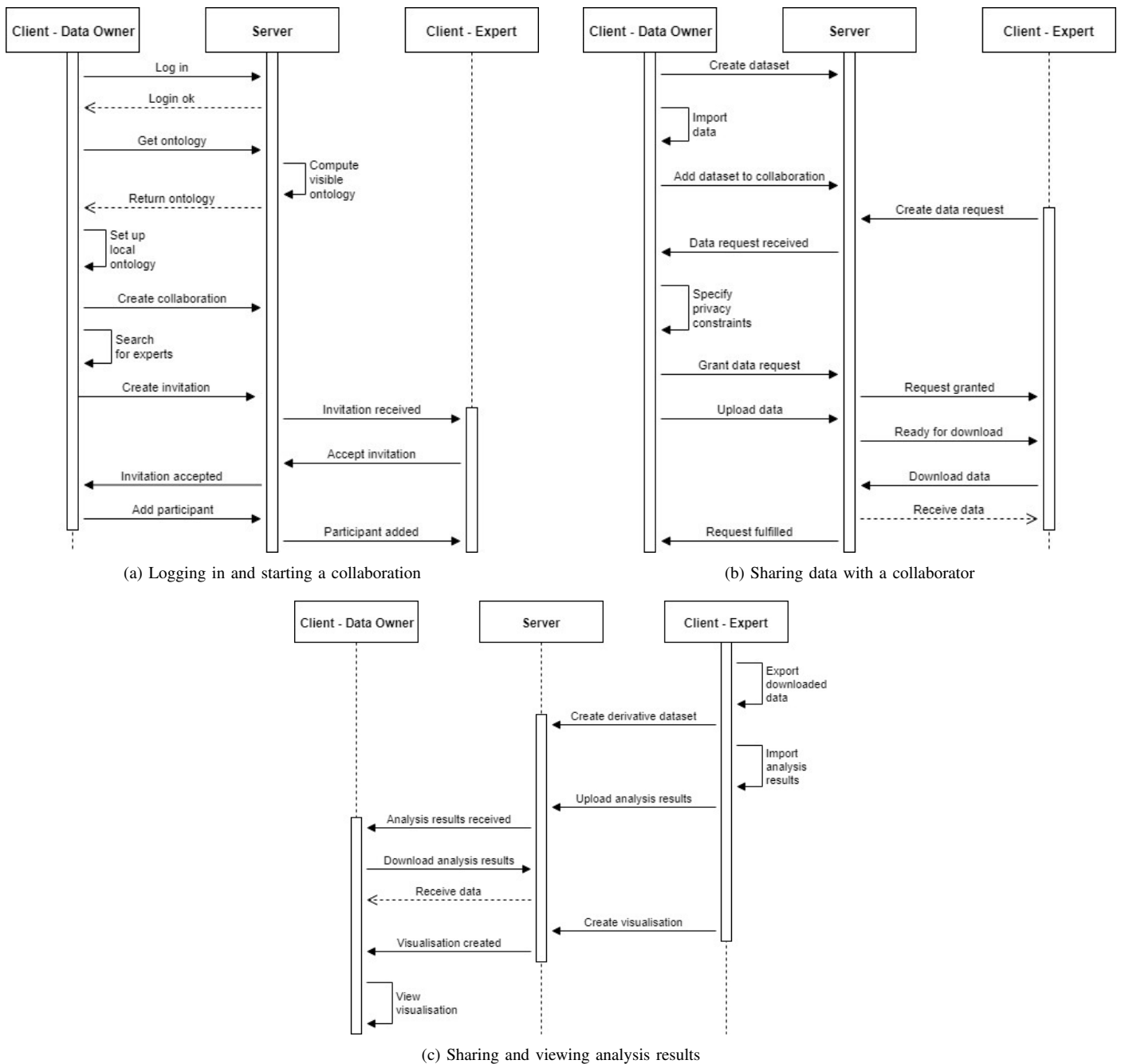


Fig. 9. Sequence diagrams summarising the events of the application scenario

experts to work collaboratively while helping non-experts understand what the experts are doing.

Besides multiple experts, it is also conceivable that there could be multiple data owners participating in the same collaboration. So far we have been implicitly assuming a collaboration involving just one data owner and one expert, but usage will show whether this is in fact a representative example of how these collaborations are carried out. The more participants there are in a collaboration, and the more diverse their roles and backgrounds, the more complex the process will be, but arguably the major tasks to be performed by the

collaborators are still the same as in the simplest case, and therefore the requirements of the collaboration platform are not necessarily drastically affected.

Even with an effective solution to the problem of facilitating participation of non-experts in data analysis activities, the principal role of data owners in collaborative knowledge discovery will be to drive the specification of objectives and constraints through negotiation. Supporting this part of the collaboration process will therefore continue to be one of our primary interests in our future work on the collaboration platform. In fact, the focus of the process in general is likely

to shift toward this part in the future, as advances in artificial intelligence make it increasingly feasible to automatically determine how to achieve knowledge discovery objectives once humans have determined what those objectives are.

Regardless of the outstanding research problems, the application scenario demonstrates that the software platform we have developed can already be used to carry out a collaboration that potentially increases the value of self-tracking data for a data owner. The motivation of the expert is currently not taken into consideration, but this is also something we will address in future work. In the immediate future, we are validating the current implementation and gathering data to inform development decisions by means of a trial, similar to the application scenario, involving a number of test users interested in gaining insights into their sleep patterns through data collection and collaborative analysis.

VI. CONCLUSION

In this paper we discussed collaborative knowledge discovery as a means of extracting additional value from self-tracking data, such as activity or sleep data. To support such collaborations, we have developed an online software platform to facilitate all of the major tasks of the collaboration process, which we have identified as finding collaborators, negotiating terms, obtaining resources, performing analyses and evaluating results. We have translated these tasks into software requirements and implemented these in a proof-of-concept system. Although the functionality of the current implementation is limited, it can already be used to carry out potentially beneficial real-world collaborations, which we demonstrated here by presenting an application scenario involving periodicity analysis of sleep data. Several topics requiring further research have been identified and will be addressed in future work.

ACKNOWLEDGMENT

Lauri Tuovinen is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement number 746837. Alan Smeaton is part-funded by Science Foundation Ireland under grant number SFI/12/RC/2289 (Insight Centre).

REFERENCES

[1] L. Tuovinen and A.F. Smeaton, "Unlocking the black box of wearable intelligence: ethical considerations and social impact", in *Proc. 2019 IEEE Congress on Evolutionary Computation*, 2019, pp. 3235-3243.

[2] L. Tuovinen and A.F. Smeaton, "Ontology-based negotiation and enforcement of privacy constraints in collaborative knowledge discovery", *presentation at the 2nd International Workshop on Personal Analytics and Privacy*, 2018, Web: http://kdd.di.unito.it/pap2018/papers/PAP_2018_paper_2.pdf, accessed 3 Sept., 2019.

[3] L. Tuovinen and A.F. Smeaton, "A domain ontology and software platform for collaborative personal data analytics", in *Proc. 16th International Conference on Cooperative Design, Visualization, and Engineering*, 2019, pp. 1-10.

[4] F. Hu, A.F. Smeaton and E. Newman, "Periodicity detection in lifelog data with missing and irregularly sampled data", in *Proc. 2014 IEEE International Conference on Bioinformatics and Biomedicine*, 2014, pp. 16-23.

[5] M. Erdmann, R. Fischer, C. Glaser, D. Klingebiel, M. Komm and others, "A web-based development environment for collaborative data analysis", *Journal of Physics: Conference Series*, vol. 523, no. 1, 2014, article 012021.

[6] N. Karacapilidis, S. Christodoulou, M. Tzagarakis G. Tsiliki and C. Pappis, "Strengthening collaborative data analysis and decision making in web communities", in *Proc. 23rd International Conference on World Wide Web*, 2014, pp. 1005-1010.

[7] E. Afgan, D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier and others, "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update", *Nucleic Acids Research*, vol. 44, no. W1, 2016, pp. W3-W10.

[8] R. Marée, L. Rollus, B. Stévens, R. Hoyoux, G. Louppe and others, "Collaborative analysis of multi-gigapixel imaging data using Cytomine", *Bioinformatics*, vol. 32, no. 9, 2016, pp. 1395-1401.

[9] C. Diamantini, D. Potena and E. Storti, "A semantic-aided designer for knowledge discovery", in *Proc. 2011 International Conference on Collaboration Technologies and Systems*, 2011, pp. 86-93.

[10] M. Deshpande, D. Ray, S. Dixit and A. Agasti, "ShareInsights - an unified approach to full-stack data processing", in *Proc. 2015 ACM SIGMOD International Conference on Management of Data*, 2015, pp. 1925-1940.

[11] E. Kandogan, M. Roth, P. Schwarz, J. Hui, I. Terrizzano and others, "LabBook: metadata-driven social collaborative data analysis", in *Proc. 2015 IEEE International Conference on Big Data*, 2015, pp. 431-440.

[12] Apache Zeppelin website, Web: <https://zeppelin.apache.org>, accessed 11 Sept., 2019.

[13] Oura website, Web: <https://ouraring.com/>, accessed 11 Sept., 2019.

[14] Spyder website, Web: <https://www.spyder-ide.org/>, accessed 13 Sept., 2019.

[15] SciPy website, Web: <https://scipy.org/index.html>, accessed 13 Sept., 2019.

[16] MyData Global website, Web: <https://mydata.org/>, accessed 14 Oct., 2019.

[17] European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)", 2016.