# Exploring Correlational Information in Aggregated Quantified Self Data Dashboards

**Simon L. Jones**
University of Bath
Department of Computer Science
Bath, UK
s.l.jones@bath.ac.uk

## Abstract
Data aggregation platforms are often depicted as a panacea for users wanting to examine correlations within the multi-faceted data that they collect. In this paper we describe inherent challenges with the provision of multi-faceted, correlational information in data aggregation tools, and present a set of hypotheses related to these challenges. We point to design considerations for improving such tools and describe an on-going study of one such tool, Exist.io, in which we aim to explore the issues discussed.

## Author Keywords
Quantified self; Data dashboards; Data aggregation; Mashups; Correlation Analysis; Self tracking.

## ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

## Introduction
The practice of self-tracking has become increasingly popular in recent years through the widespread use of sensor-enriched smart devices (e.g. smartphones and smartwatches), improved biometric sensors (e.g. heart-rate trackers, body mass index scales), and services

| Connected Service | Attribute (Variable) |
|---|---|
| Exist for Android | Mood score, mood note |
| Fitbit, Jawbone UP, Withings | Steps, sleep, weight |
| Google Calendar, iCloud, iCal | Events, time spent in events |
| RescueTime | Productive time, neutral time, distracting time |
| Forecast.io | Weather conditions |
| Last.fm | Tracks played |
| Twitter | Tweets, mentions |
| Instagram | Number of posts, comments, likes |
| Swarm (Foursquare) | Check-ins, location |

**Table 1**: Services that can be connected to Exist.io in order to aggregate quantified self data. Some services provide more than one data attribute/variable, all of which can be included in the automated correlation analysis.

specifically designed for monitoring purposes (e.g. RescueTime[1]). Previous research has reported that many systems for personal data capture focus on a single facet of people's lives (e.g. only capturing physical activity), whilst claiming that the benefits of such systems are likely to increase if they can offer support for associating multiple life facets [e.g. 7]. Moreover, it is commonly suggested within the quantified self (QS) research community that tracking is most interesting and useful when it takes a holistic view of a users' life [4]. One of the most common types of insight that self-trackers have sought to acquire is an understanding of the correlations between different aspects of their behavior, for example; learning how their sleep impacts on their productivity[2], how their physical activity affects their mood[3], or how their stress levels are associated with their social media use [3].

To this aim, we are beginning to see the emergence of web-based aggregation platforms for quantified self data (e.g. Exist.io[4], Tictrac[5], Zenobase[6]), which combine tracked data from multiple sources associated with many different life facets and support the investigation of relationships between variables by providing visualizations and automated analysis of the data. The Exist.io platform provides explicit support for exploring correlations (see Figure 1 and 2) and enables its' users to connect various services for tracking user attributes and behaviours. Table 1 provides a list of the services and attributes currently available in Exist.io. All data provided across these connections are analyzed continuously in order to identify correlations between attributes. Information about the strongest positive and negative correlations that are detected are presented within Exists' own data dashboard as graphs of the correlated variables over time and readable explanations of the correlation. Figure 1 illustrates several examples of such correlations and Figure 2 shows the detailed graph of one particular correlation.

Hadadi et al. suggest that the ability to relate data across different facets, for example "correlating physical activity with other data such as calorie intake or mood" is likely to result in more "appealing inferences" for users, and increase engagement in collecting and using personal data [6]. Although data aggregation, or 'mashup', tools have been explored in previous research (e.g. the Mobile Health Mashup [12]) there remain a number of open research questions and design challenges associated with this type of tool. For example, several researchers question the best way to represent correlations to users, particularly "typical consumers" [7] and those less familiar with data analysis [1, 8]. Others highlight the problem with inaccuracies in tracking data affecting the reliability of the correlational insights that are provided [13].

In this workshop paper we describe several additional research areas to explore with respect to providing users with correlational information in aggregated data tools, namely; dealing with high-dimensional data, moving beyond 'stating the obvious', and assisting users to act on correlational information. We then

[1] http://www.rescuetime.com

[2] https://forum.quantifiedself.com/thread-personal-experiments-qs-study-relating-to-sleep

[3] https://www.ginsberg.io/activity-mood-diary-app/

[4] http://www.exist.io

[5] http://www.tictrac.com

[6] https://zenobase.com

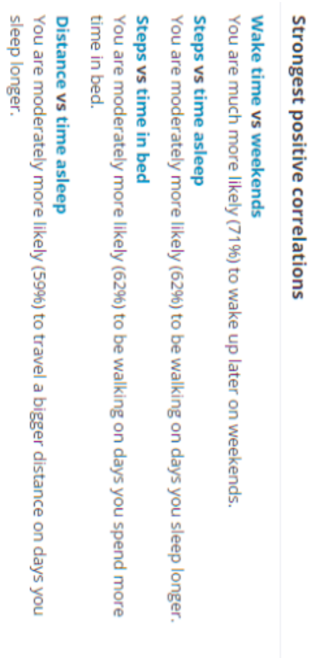


**Figure 1**: An example of Exist.io's[4] list of strongest positive correlations (https://exist.io/)

describe an on-going study that aims to explore these issues and inform the design of data aggregation tools for QS.

## Dealing with high-dimensional data

One of the significant challenges that data aggregation tools are likely to face is that many interesting insights and correlations may be lost in the vastness of the accumulated data. For example, a tool that explores relationships between 20 different variables has the potential to report up to 190 correlations. Previous research has frequently reported that users of quantified self technologies cite problems related to "drowning in data" [9] and having insufficient time to go through the results [*7*]. Hence, we are interested in understanding the requirements for automatically detecting interesting insights from within a user's data, such that the burden of exploration is driven by the system, rather than the responsibility of the user.

A common pitfall identified in the design of information dashboards is that of ineffective highlighting of important information [5]. It is argued that an effective information dashboard should immediately draw the user to the information that is most important and that requires immediate attention [5]. In the context of data aggregation tools aimed to help users "make sense of their lives"[7], it is not necessarily obvious what information a particular user is likely to consider 'important'. At present, services such as Exist.io place their focus on reporting the *presence* of positive and negative correlations (see Figure 1). However, one challenging aspect of automatically filtering and highlighting correlational information is that the *absence* of a strong correlation may often be just as interesting or useful as its presence. For example, an individual learning that his/her sleep and cognitive performance are *not* correlated could prove to be an intriguing and useful insight.

[7] https://zenobase.com

To begin investigating the problem of automatically selecting which information to convey to a user, we present a number of hypotheses related to the utility of correlational information. Our first hypothesis (H1) is that comparisons between: a) correlations within actual data and b) those predicted by a user, based on their mental model of their behaviour, are useful for information filtering. We anticipate that many of the beneficial insights gained from the use of data aggregation platforms will either be associated with considerable discrepancies between expected results (based on the user's mental model of their behaviour) and those obtained, or reliable evidence of correlations that are already suspected. Evidence of quantified-selfers reporting insights related to both confirmation and contradiction of existing knowledge are revealed in [1]. This hypothesis also raises the question of how a user might go about providing a representation of their mental model to the system, such that these discrepancies and similarities can be detected.

Another possible mechanism for selecting and filtering information could be to utilize feedback from other users about the utility of the correlations provided. In other words, if users are able to rate the utility of correlational information that is provided to them, this feedback could be used to drive information presentation for similar users. Although specific correlations are likely to be different across individuals, future research should seek to identify the factors that

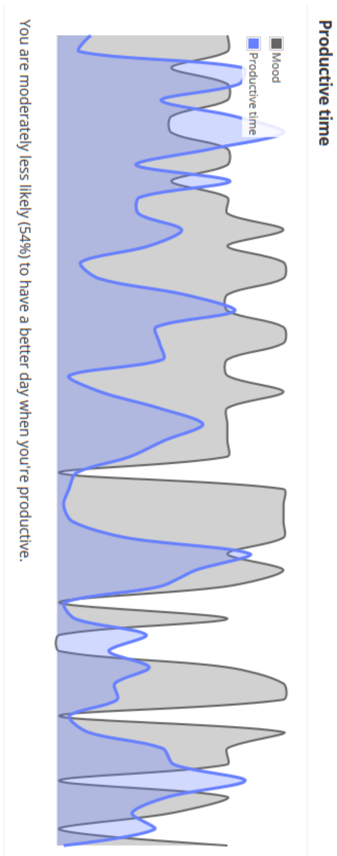


**Figure 2**: A graph visualization of the correlation between Mood and Productivity in Exist.io[4] (https://exist.io/)

influence their perceived utility. One could conceive that there might be rules for predicting the utility of correlational information, despite these individual differences.

Our second hypothesis (H2) is that comparisons between a) a user's correlations and b) correlations of other users will also influence the perceived utility of the insight. For example, it may be interesting to explore whether users value information that uncovers unique characteristics of their behaviour, when compared with their peers, or alternatively, there may be value in simply learning that their behaviour is 'typical' when compared to others.

Our third hypothesis (H3) is that some correlations are useful by virtue of being difficult to predict. We anticipate that some people may struggle to foresee what their aggregated data is likely to reveal about them, and would therefore be interested to find out the true nature of certain correlations. For example, a user may have no idea whether her interactions on social media are correlated with her mood (either due to increased use of social media when feeling happy or sad, or owing to her mood being directly affected by engagement with social media). Future research could explore the factors that influence a users' ability to make accurate predictions about the correlations between variables.

In addition, we suggest that it is important to explore whether the utility of correlational information can be predicted by quantifiable characteristics of the user, such as their goals for engaging in self-tracking [2], or their individual values (e.g. whether they are motivated by personal success, conformity, stability of self, curiosity, etc. [10]), such that a system can tailor the presentation of information to a particular user.

## Moving beyond stating the obvious

Our examination of services such as Exist.io has also revealed that not all of the correlations that surface are likely to be considered insightful. For example, many of the strongest correlations are inevitable, due to interrelatedness of the variables being measured, e.g. "You are certain (100%) to travel a further distance on days you take more steps" and "You are almost certain (86%) to be active when you take more steps". Other correlations may also be considered 'obvious' because they reflect typical characteristics of human behaviour, for example, being more productive on weekdays and less productive at weekends, or being happier when the weather is better.

Whilst it has been suggested that multi-faceted tracking tools are likely to increase engagement in collecting and using data [6], it may be possible that they have the opposite effect when presenting information that is considered to be obvious. Anecdotally, we have observed users of such services commenting on the obviousness of the conclusions provided (e.g. see Figure 3). The developers/designers of these services often suggest that many of most insightful correlations take time to surface (e.g. see Figure 3). We therefore suggest that future research should seek to understand the impact that the presentation of such correlations have on users' perceptions and engagement with the service.

## Acting on Correlational Information

Li et al. [7] reported that for many users of self tracking technologies, a key problem is "not knowing

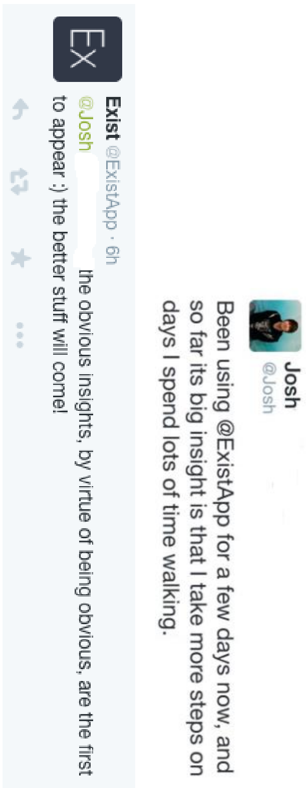


**Figure 3:** An Exist.io user commenting on Twitter about the obviousness of the correlations being presented.

what to do with the information provided", and that suggestions for how to take action are often needed. As data aggregation services become more mainstream we anticipate that they are increasingly likely to attract users who do not have a clearly defined problem to address, or a hard-set end goal for using the system. One of the common activities for members of the quanitifed self community is that of 'self-experimentation': using preliminary analysis to generate hypotheses that can be followed up with more rigourous testing. We expect that self-experimentation may be a likely action for users of data aggregation services when it is unclear whether the relationship between two variables is due to correlation or causation. For example, a user presented with the insight that he is most productive when he has events in his online calendar may raise the question of whether the act of planning his time in advance, and thus creating calendar entries, has enabled him to work more effectively throughout the day, or whether this correlation exists simply because his productivity is highest on working days, and his calendar entries are typically for working days also.

We aim to investigate whether platforms such as Exist.io encourage users outside of QS special interest groups to engage in self-experimentation in order understand their behaviour more deeply. We believe that understanding the type of questions that correlations provoke and the types of action that a user may wish to take as a result, is useful for informing the design of data aggregation tools. For example, in order to support users, a platform might suggest ways in which the user can investigate further – either by conducting further experiments, or by interrogating the data in more detail (e.g. exploring the effect of removing working/non-working days from the correlation analysis).

## Study Overview and Future work

In order to investigate the issues discussed and test the hypotheses offered in this paper we are currently conducting a study, in which participants with varying experiences of self-tracking, are invited to use the Exist.io platform for a period of several months. Participants will connect as many of the supported tracking services that they currently use (see Table 1), or install the required software to track behaviours that they wish to include (e.g. RescueTime to track laptop and smartphone use, Last.fm to track music listening). Participants will be asked to provide predictions about correlations in their data, based on their understanding of their behavior, enabling us to explore the relationship between their mental models and the perceived utility of correlational information (H1). Participants will also be asked whether they think the service is likely to provide useful insights, such that we can examine whether users are able to make reasonable predictions about the utility that data aggregation tools provide, or whether there are unexpected benefits.

After an initial period of accumulating data, users will be given access to their 'Correlations' page within the Exist.io dashboard. Many of the correlational analyses require a minimum of 30 days-worth of data before results are calculated. Participants will then be interviewed about these correlations to determine what they have learnt about their behaviour from the reported correlations, and asked about their perceptions of the service and its utility, in order to shed light on the issues discussed within this paper. The collected data and feedback from users will support our investigation about whether we can develop mechanisms to automatically select the most useful/interesting insights to present to users data from the vast amount of data collected (H2 & H3). Participants will then continue to use the Exist platform for a further period in order for us to understand whether the utility of the service changes over time and how the correlational information provided affects participants engagement with the service. We anticipate that the findings from this study will provide valuable information related to the issues discussed in this workshop paper, and provide useful design implications for QS data aggregation services.


## Acknowledgements
We thank Edward Bradley for his contribution to discussions about this work.


## References

1. Eun Kyoung Choe, Bongshin Lee, and M. C. Schraefel. (2015). Revealing Visualization Insights from Quantified-Selfers' Personal Data Presentations, Computer Graphics and Applications, IEEE , vol.PP, no.99, pp.1,1
2. Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. (2014, April). Understanding quantified-selfers' practices in collecting and exploring personal data. In Proc. 32nd annual ACM conference on Human factors in computing systems (pp. 1143-1152).
3. Emily Collins, Jon Bird, Anna L. Cox, and Daniel Harrison. (2014). Social networking use and RescueTime: The issue of engagement. In Proc. 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. (pp. 687 - 690).
4. Tilman Dingler, Alireza Sahami and Niels Henze. (2014). There is More to Well-being than Health Data: Holistic Lifelogging through Memory Capture. CHI 2014 - Workshop Beyond Quantified Self: Data for Wellbeing
5. Stephen Few. (2006). “Common Pitfalls in Dashboard Design.” Perceptual Edge.
6. Hamed Haddadi and Ian Brown. (2014). Quantified self and the privacy challenge. Technology Law Futures.
7. Ian Li, Anind Dey, and Jodi Forlizzi. (2010, April). A stage-based model of personal informatics systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 557-566). ACM.
8. Ian Li, Jon Froehlich, Jakob E. Larsen, Catherine Grevet, and Ernesto Ramirez.. (2013). Personal informatics in the wild: hacking habits for health & happiness. InCHI'13 Extended Abstracts on Human Factors in Computing Systems (pp. 3179-3182). ACM.
9. Ian Li. (2015). Beyond Reflecting on Personal Data: Predictive Personal Informatics. CHI 2015 Workshop - Beyond Personal Informatics: Designing for Experiences with Data.
10. Shalom H. Schwartz. (2012). An overview of the Schwartz theory of basic values. Online Readings in Psychology and Culture, 2(1), 11.
11. Christopher Trickler. (2013). An overview of self-monitoring systems. In Proc. Southern Association for Information Systems Conference (pp. 197-202).
12. Konrad Tollmar, Frank Bentley, Cristobal Viedma, and I. L. Libertyville. (2012). "Mobile health mashups." *Proc. PervasiveHealth* 12.
13. Hongbin Zhuang. (2014). Was I productive? Examining the reliability of the Quantified Self technology. Faculty of Brain Sciences, University College London.