

*This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.*

**Please cite the original version:** Santonen, T. (2022) Novel approach to collect longitudinal open data with students. In Leandro Bitetti, Iain Bitran, Steffen Conn, Jessica Fishburn, Eelko Huizingh, Marko Torkkeli, Jialei Yang (Eds.) Proceedings of the XXXIII ISPIM Innovation Conference.

---

## Novel approach to collect longitudinal open data with students

---

Teemu Santonen

Laurea University of Applied Sciences, Vanha maantie 9, 02650,  
Espoo, Finland  
E-mail: teemu.santonen@laurea.fi

**Abstract:** This study proposes a novel crowdsourcing based operating method for universities to collect longitudinal data as a part of degree programs. The perceived benefits of such a system for the teachers and students as well as the key implementation obstacles and mitigate possibilities are evaluated among key personnel (N=17) of one case university. The proposed operating method is introduced and compared against the crowdsourcing definition. The perceived benefits were related to (1) improved pedagogics, (2) intensified collaboration between university, students and working life, (3) accumulation of new research skills and (4) novel publication opportunities. The key challenges included (1) legal, licence and ethical issues, (2) concerns relating lack of competence, (3) technical systems compatibility issues, (4) lack of resources and motivation among teachers and students and (5) continuity problems in cooperation and data collection.

**Keywords:** Longitudinal data, Crowdsourcing, Pedagogical innovation, Open science, Open data, Open innovation

---

### 1 Introduction

It is surprising how few innovation studies have adopted longitudinal methods (Perks and Roberts, 2013). Furthermore, even if longitudinal studies have been widely published since 1970s, longitudinal qualitative studies have gained main attention while qualitative or mixed-method have been less popular (Hassett and Paavilainen-Mäntymäki, 2013). Longitudinal data is important since some phenomena can only be evaluated in a longitudinal research setting. Furthermore, the user-centered innovation and design - a process where products or services are co-created with the users - have also gained popularity during the past decade (Norman and Draper, 1986). One-off quantitative studies often iron out the details, which user-centered innovation practices emphasis. On the other hand the key disadvantages of the longitudinal studies includes long implementation time and expensiveness, making it difficult for researchers and companies to execute them especially when following qualitative research approach. However, developing new product, services and processes, and evaluating their impact calls out in-depth understanding people's attitudes and needs, which only a qualitative research approach can deliver.

Lately user-centered (also known as human-centered) innovation grounded on design thinking have gained increased popularity among the innovation scholars and practitioners. According to (Lockwood, 2009) design thinking is "a human-centered innovation process

that emphasizes observation, collaboration, fast learning, visualization of ideas, rapid concept prototyping, and concurrent business analysis” whereas user-centric innovation as a term is describing a process in which end-users can influence how a design (or innovation) takes shape (Norman and Draper, 1986). However, due lack of resources especially SMEs are forced to focus on daily business, instead of implementing resource-intensive user-centric innovation and design thinking approaches (Acklin, 2010; Fischer et al. 2019). Therefore, this study proposes a novel operating method in degree education in which longitudinal research data can be systematically collect as a part of course studying task at the university level. In design thinking, the innovation process is typically divide into (1) exploration of the problem space, (2) exploration to the solution space, and (3) the iterative alignment of these two phases (Lindberg et al. 2011). The proposed operating method focuses on the double diamond discover-phase (Council, 2015), thus covering the first part of the exploration of the problem space. The discover phase aims to understand the existing user needs, gain insights about the current problem space and situation in order to formulate the key challenges which needs to be solved.

### *Objectives of this study*

After introducing the operation model, this study aims to answer to the following research questions. What are the main benefits for teachers (RQ1 and students (RQ2), if the described longitudinal research data collection model is implemented at higher education institute (later HEI) level (i.e. data collection activities are covering all degree programs and multiple courses). What are the key obstacles preventing the successful implementation of such a strategy (RQ3) and how to mitigate or remove these obstacles (RQ4)? This study is structured as follows. *First*, interlinkage to related theoretical concepts are made. *Second*, research design is presented including an overview of proposed operating model, sample selection, data collection and analysis procedures. *Third*, the key finding regarding RQ1 to RQ4 are summarized. *Finally*, conclusions are made.

## **2 Theoretical foundations of the proposed operating method for longitudinal research data collection**

Single HEI – Laurea University of Applied Sciences – from Finland was selected as a sample case. The sample HEI has recently renewed its strategy and made a strategic choice to adopt a new operating method in which the longitudinal research data collection process is integrated as a part of all degree programs (Laurea, 2022). According to strategy, the main aim is to create a special strength by systematically collecting longitudinal research data in all areas of expertise to produce unique, open research material for the use of society at large. In practice, information resources, scientific data and the curriculum objectives are used to formulate learning, internship and thesis assignments that match the students’ level of competence and can help expand existing data repositories with the new data. Assignments can contribute to data repositories e.g. with qualitative observations or interviews, and later on the augmented data repositories and the resulting cumulative data can be used in theses, in participatory research and regional development as well as in service business, in compliance with the FAIR (Findable, Accessible, Interoperable and Reusable) Data principles (Wilkinson et al. 2016).

According to operational guidelines (Bingham, 2021), the starting point for the longitudinal data collection can be (1) course learning objective, (2) objective derived from the university research program or (3) an observed need in one of the university expertise areas. Once the data collection plan is structured, a person responsible for the research plan will fill a form for the decision making process. The form includes the following key information regarding the planned data collection: (1) basic information about the data and degree course where the data is collected, (2) research plan, (3) data management plan, (4) comments on ethical issues, personal data processing and external partners. Based on the provided information, the degree program development team will make go/no-go decision. After the approval decision, data collection will be carried out on all implementations of the course in question. The students will transfer the rights to collected data by signing copyright transfer agreement. In the case of compulsory courses for those students who are not willing sign copyright transfer agreement, an alternative studying tasks if offered. If the student has an idea for the data collection, he or she can approach the research program directors, teacher who is responsible for the course or degree program.

#### *Crowdsourcing as a foundation for the proposed operating method*

The crowdsourcing term was popularized by Howe (2008). Crowdsourcing is a process where a task or tasks are delegated to a large group of people (i.e. crowd) who complete the task based on the given guidelines. According to Estellés-Arolas and González-Ladrón-de-Guevara (2012) the following eight key characteristics defines crowdsourcing. *First*, there must be a clearly defined crowd, which in this case is university students studying in a university who is implementing the proposed operating method. *Second*, there is a task with a clear goal. The tasks in this particular case is a longitudinal data collection task, which will be done as a part of certain degree course over time. *Third*, the recompense for the crowd must be clear, which in his case is the study credits once the whole course is completed. *Fourth*, the crowdsourcer must be clearly identified. At the organizational level the crowd--sourcer is the university while the crowdsourcing manager in practice is the teacher who carries out the learning task. *Fifth*, the compensation for the crowdsourcer is clearly defined, which in this case is the accumulated data. *Sixth*, an online assigned process must be participative type. In case of the proposed operating model, there is conscious participation among the student (i.e. they have enrolled in the course) and they are seeking the same end goal (i.e. collecting the data to complete the studying task). *Seventh*, an open call of variable extent is utilized, which in this case is the course enrolment process and studying task description in a pedagogical platform. *Finally*, the internet must be used. The utilization of internet is two folded. Both the enrolment and studying tasks delivery will utilize internet. As a results, it is argued that the proposed operating method can be considered as a crowdsourcing. More precisely, it is micro-tasking approach (Geiger et al. 2012) where larger job (data collection) is splitted into small chunks of work (individual studying task for one student) and then combing the resulting individual contributions into a collective result (longitudinal data set).

Open Science, open data and open innovation are disruptive phenomenon, which have gained popularity especially in Europe (Chesbrough, 2003, Vicente-Saez, and Martinez-Fuentes, 2018). Governments and European commission are pushing universities to become more open, thus forcing them to seek novel approaches to foster open science and open data movement (e.g. Ramjoué, 2015, Forsström et al. 2020). University students are underutilized resource when it comes to large scale big data collection, which requires

manpower (Santonen, 2012). A recent study systematically reviewed crowdsourcing practices for education (Jiang et al. 2018). However, crowdsourcing in educational context appeared to omit micro tasking, which has gained popularity among research communities to lower research expenses (Difallah et al. 2015; Mason and Suri 2012).

In contrast to Amazon's Mechanical Turk (i.e. popular microtasking platform) learning activities in degree programs are subordinated to curriculum specifications. Thus, teachers are obligated to set up learning activities that support desired learning outcomes as defined in the curriculum objectives. Modern pedagogical approach argue that teaching and learning should take place in a holistic system, which goes beyond classroom and university premises. Trialogical learning approach is providing a theoretical framework for the longitudinal data collection concept aiming to produce open qualitative and quantitative data for the innovation management purposes (Paavola and Hakkarainen, 2005). The key characteristics of trialogical learning are (1) shared objects of activity, (2) sustained and longstanding pursuit of knowledge advancement, (3) mediated interaction between individual and collective knowledge-creation activities, (4) cross-fertilization of knowledge practices between educational, professional, and research communities, (5) technology mediation for supporting collaborative knowledge creation and sharing and (6) development through transformation and reflection.

In summary, the proposed the operation model is grounded on the following idea. *First*, from each degree program within a university, group of courses are selected to conduct a data collection task. *Second*, the task is defined to be fit with the course objectives and the students' level of competence in order to support the learning objectives achievement. *Third*, the process is recurred for all matching courses in the coming years. *Fourth*, the collected open data will be made available via online data repository in compliance with the FAIR (Findable, Accessible, Interoperable and Reusable) data principles.

### 3 Research design

#### *Data collection*

A multi-method research approach consisting interviews, survey and focus group discussion was applied as follows (Brewer and Hunter, 2006). The key personnel (N=17) from the sample HEI were identified including (1) HEI top management (N=3), (2) persons leading degree programs or being a member of pedagogical development group (=10), (3) persons being responsible for HEI level research programs (N=3) and (4) information specialist (N=1). *First*, four in-depth interviews were conducted to gain the overall understanding about the key research questions and make sure that all vital viewpoints were considered in the research setting. Duration of the each interview was ca. one hour. *Second*, the identified key personnel were asked to fill open-ended online survey including questions relating RQ1- RQ4. Each responded were asked to think about the situation from their own role point of view (e.g. top management, research program leader, information specialist and degree programs leader / pedagogical development group member). The received qualitative data from interviews and open-ended survey questionnaire were analyzed by following thematic analysis six phases approach proposed by (Braun and Clarke, 2006). *Third*, a focus group discussion session (ca. 1.5 hours) was arranged to

discuss the findings of the survey with the respondents in order to validate and refine the survey findings (Nyumba et al. 2018).

## 4 Results

### *Perceived benefits for teachers and students*

The perceived benefits for teachers and students regarding crowdsourcing based longitudinal open data collection are compared in Table 1.

**Table 1** The perceived benefits for teachers and students regarding crowdsourcing based longitudinal open data collection

<i>Benefits</i>	<i>For teachers</i>	<i>For students</i>
New pedagogical materials and improved pedagogics	Getting new data and case examples for pedagogical materials with less work. Getting support to course development and student guidance due better understanding of local and regional needs and challenges.	Up-to-date working life data driven educational materials grounded on collected data. Collaboration with working life actors during the course.
Research skills development	Improving one's research and methodological skills regarding managing open data collection projects grounded on crowdsourcing.	Improving one's research and methodological skills relating open data collection, analysis and utilization due practical experience.
Knowledge and competence gap identification	Identifying competence and knowledge gaps due better understanding of current working life, market and customer needs.	Gaining better understanding and awareness of customer and working life needs. Gaining understanding on topical business challenges which might helping one's career planning.
Publication opportunities	Increases possibilities to do research publications based on collected data.	Ability to do thesis without itself collecting the data or getting richer and larger data for one's thesis.
Networking and collaboration opportunities with working life	Intensified collaboration and networking with local and regional working life stakeholders. Intensified opportunities for regional development.	Gaining opportunities to acquire internship and contract based thesis from working life.
Improved possibilities to gain research funding.	Getting support to funding application preparation due better understanding the current-business challenges and customer needs.	Increased opportunities to get contract based thesis from working life.

### *The key challenges and mitigation approaches*

The key challenges and suggested mitigation approaches are presented in Table 2.

**Table 2** The key challenges regarding data collection, data utilization and mitigation approaches

---

<i>Key challenges</i>	<i>Mitigation approaches</i>
Issues regarding legal, contract, licence and permissions including such as consent from study participants and contracts with organizations where the data is collected. Agreement on data ownership and usage rights between university and students while enabling students to carry out compulsory courses without participating the data collection.	Defining common templates for consents and legal contracts. Open and transparent communication regarding the benefits of longitudinal data collection for different stakeholders.
Research ethical issues such as concerns related students and teachers competences, skills and knowledge regarding research ethics, data protection, consent, privacy and personal data processing issues especially in the case of health data.	Providing clear ethical guidelines, introducing quality control mechanisms and predefining what kind of data can be securely collected.
Data collection processes and technical systems compatibility issues with course objectives, pedagogical platforms including interplay between pedagogical platforms and data collections tools in all data management phases.	Investing on a proper technical platform and developing tailor-made solutions if needed to enable interplay between different systems.
Lack of teacher resources to carry out and manage data collection tasks due increased workload and unclear roles.	Ensuring enough working time for teachers especially when new model is introduced. Clearly defining who is responsible for different phases (planning, collection, cleaning, analysis and storage)
Concerns regarding teacher's and student's ability to carry out high quality data collection and data management.	Defining easy to use support documentation and operational process. Providing training for teachers and students by e.g. introducing training badge to verify minimum skills set. Adjusting data collection task according to students' competence level. Establishing support team, who provides assistance and consultation for the teachers
Continuity problems in cooperation and data collection with the external stakeholders due (1) changing working life needs, (2) changing course contents and (3) annual curriculum variations.	Critically evaluating and agreeing what kind of data is possible to collect in longer term.
Students, teachers and partners lack of motivation if they do not feel that data collection is meaningful or it does not provide any benefits. Risk is high especially when operational model is starting and there is no prior data yet.	Clarifying and communicating the data collection purpose and goal. Allowing teachers, students and partners to influence and choose what kind of data is collected. Selecting teachers who has strong personal research interest and ensuring that data collection has clear linkage to one's work.

---

## 5 Discussion and conclusion

This study introduced a novel pedagogical model for universities to collect open longitudinal data as a part of degree program courses. The perceived benefits, key obstacles and obstacle mitigation activities of the suggested concept were identified from teachers and student point of view. Based on the obtained results, the concept has multiple potential benefits covered following topics: (1) improved pedagogics and new pedagogical materials, (2) improving both teachers and students research and methodological skills, (3) helping to identify knowledge gaps, (4) intensifying collaboration with working life actors, (5) gaining opportunities for more impactful regional and local development, (5) opportunities to gain research funding due to better understanding of current business and changing business and customer needs and (6) having new opportunities for publishing.

The key concerns covered legal, contract, licence and permissions issues, which are mandatory to settle before starting the data collection. Especially the data ownership and copyright issues can be difficult to agree with the students especially if the benefits remain unclear for the students. Therefore, a need for open and transparent communication regarding the benefits of longitudinal data collection for different stakeholders is highlighted. Since students and teachers skills regarding research ethics varies greatly (e.g. between first year and final year students) it is important to define data collection tasks according to students' competence level. Special attention should be dedicated to health data collection, since unintentional mistakes can happen when there is a lot of secondary information available, which can then lead to persons identification. Technical systems compatibility issues are also expected, since pedagogical platforms are not primarily designed for observation or interview data collection. Therefore, it is likely that the integration between data collection, course tasks, data cleaning and data storage will need multiple technical platforms. When data collection volume becomes high, this might lead to serious difficulties. Resistance among teachers and students is expected. Teachers are primarily afraid of the extra workload on top of the normal course activities, since implementing new pedagogical approaches will always take more time than carrying out the proven teaching approaches. A need to provide strong support for teachers, who can then help students, is clearly needed when starting the implementation. Ready-made guidelines and training courses must be provided to lower the bar for participation. Furthermore, open, transparent and active communication regarding the benefits of longitudinal data collection for different stakeholders is expected to help increase the motivation. Allowing teachers, students and partners to influence and choose what kind of data is collected is also perceived to have a positive impact on willingness to participate. Continuity problems in cooperation and data collection with external stakeholders is also among the key challenges. The primary idea in longitudinal data collection is to keep the questions unchanged over time. However, working life needs, course contents and curriculums are changing over time and it can be difficult to define research questions that remain relevant over a long period of time. Evidently finding the right research topics is the key challenge.

This study has the following limitations. The perceptions are grounded only on HEI personnel perceptions. Thus, further studies should focus on evaluation of students' opinions and perceptions regarding the benefits and challenges of the proposed model. Secondly, the case study university had finished the few piloting studies. As a result, the respondents' opinions were grounded only on perceptions. It is suggested that this study should



be replicated in different universities in order to evaluate if proposed concept perceived benefits and challenges are similar in different setting. Furthermore, a new study should be executed when the sample university has more practical experience regarding the actual data collection from various course settings.

## References and Notes

- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57-71). American Psychological Association.
- Acklin, C., (2010). Design-driven innovation process model. *Design Management Journal*, 5(1), pp.50-60.
- Bingham, .C: (2021) Pitkittäisdatalla vaikuttavuutta Laureaan. Process guidelines document, only accessible in Laurea intranet.
- Brewer, J. and Hunter, A., 2006. *Foundations of multimethod research: Synthesizing styles*. Sage.
- Chesbrough, H.W., (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.
- Council, D. (2015) What is the framework for innovation? Design Council's evolved Double Diamond. The Design Council. Available at: <https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond> (Accessed 29.4. 2022).
- Difallah, D.E., Catasta, M., Demartini, G., Ipeirotis, P.G. and Cudré-Mauroux, P., (2015). The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th international conference on world wide web* (pp. 238-247).
- Estellés-Arolas, E. and González-Ladrón-De-Guevara, F., (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2), pp. 189-200.
- Fischer, S., Redlich, B., Lattemann, C., Gernreich, C. and Pöppelbuß, J., (2019). Implementation of Design Thinking in an SME. In *ISPIM Conference Proceedings* (pp. 1-14). The International Society for Professional Innovation Management (ISPIM).
- Forsström, P.L., Lilja, E. and Ala-Mantila, M., (2020). *Atlas of open science and research in Finland 2019: Evaluation of openness in the activities of higher education institutions, research institutes, research-funding organisations, Finnish academic and cultural institutes abroad and learned societies and academies*. Final report.
- Geiger, D., Rosemann, M., Fieft, E. and Schader, M., (2012). Crowdsourcing information systems-definition typology, and design (pp. Paper-53). AISeL
- Hakkarainen, K. and Paavola, S., (2007). February. From monological and dialogical to triological approaches to learning. In *A paper at an international workshop" Guided Construction of Knowledge in Classrooms*.
- Hassett, M.E. and Paavilainen-Mäntymäki, E., (2013). Longitudinal research in organizations: an introduction. In *Handbook of longitudinal research methods in organisation and business studies*. Edward Elgar Publishing.
- Howe J. (2008), *Crowdsourcing: How the Power of the Crowd is Driving the Future of Business*. Business Books, Great Britain.
- Jiang, Y., Schlagwein, D. and Benatallah, B., (2018), June. A Review on Crowdsourcing for Education: State of the Art of Literature and Practice. In *PACIS*.

Laurea (2022). Strategy 2030 of Laurea University of Applied Sciences. [https://www.laurea.fi/globalassets/laurea/documents/strategia-2030\\_eng-lowres.pdf](https://www.laurea.fi/globalassets/laurea/documents/strategia-2030_eng-lowres.pdf) (Accessed 29.4. 2022).

Lindberg, T., Meinel, C. and Wagner, R., (2011). Design thinking: A fruitful concept for it development?. In *Design thinking* (pp. 3-18). Springer, Berlin, Heidelberg.

Lockwood, T., 2010. *Design thinking: Integrating innovation, customer experience, and brand value*. Simon and Schuster.

Mason, W. and Suri, S., (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44(1), pp.1-23.

Norman, D.A. and Draper, S.W., (1986). *User centered system design: New perspectives on human-computer interaction*. CRC Press.

Nyumba, T.O., Wilson, K., Derrick, C.J., Mukherjee, N. (2018). The use of focus group discussion methodology: insights from two decades of application in conservation. *Methods in Ecology and Evolution*, 9(1), pp. 20-32.

Paavola, S. and Hakkarainen, K., (2005). The knowledge creation metaphor—An emergent epistemological approach to learning. *Science & education*, 14(6), pp.535-557.

Perks, H. and Roberts, D., (2013). A review of longitudinal research in the product innovation field, with discussion of utility and conduct of sequence analysis. *Journal of Product Innovation Management*, 30(6), pp.1099-1111.

Ramjoué, C., (2015). Towards open science: The vision of the European Commission. *Information Services & Use*, 35(3), pp.167-170.

Santonen, T., (2012). *Massidea.org: a greener way to innovate*. In *Green technologies in food production and processing* (pp. 541-568). Springer, Boston, MA.

Vicente-Saez, R. and Martinez-Fuentes, C., (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of business research*, 88, pp.428-436.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), pp.1-9.