

Review

# Meta-Analysis of Data Collect Methods

<sup>1</sup>Iman Tikito, <sup>1</sup>Mohammed El Arass and <sup>1,2</sup>Nissrine Souissi

<sup>1</sup>Mohammed V University-Rabat, EMI-SIWEB Team, Rabat, Morocco

<sup>2</sup>Ecole Nationale Supérieure des Mines de Rabat, Computer Sciences Department, Rabat, Morocco

## Article history

Received: 03-04-2019

Revised: 29-07-2019

Accepted: 23-08-2019

Corresponding Author:

Iman Tikito

Mohammed V University-  
Rabat, EMI-SIWEB Team,  
Rabat, Morocco

Email: tikito.iman@gmail.com

**Abstract:** Several organizations from different sectors depend increasingly on knowledge extracted from huge volumes of data generated by different sources, such as IoT, sensors and databases. At the core of data lifecycle, data reliability, analytics, security, scalability and use are important concerns. Coping with these issues in handling data requires understanding the challenges associated with it. Analysis process and storage devices have been widely studied. However, very few studies have explored the collect data phase. In this study we aim to analyse more the collect phase of data lifecycle to provide an optimized and smart approach. This paper aim to provide the right method to follow in data collect phase within different domain according to client needs and requirements. It provides not only a detailed view of the main steps, but also based on a prior literature review on different existing methods. This allowed us subsequently to establish a correspondence with the SLR method on which we based our method. We use an explicit example to illustrate the steps of our method.

**Keywords:** Data Lifecycle, Data Collect, SLR, Method, Big Data

## Introduction

To improve the process based on extracting knowledge from raw data, many researches in different fields faced big challenges due to multiple reasons as mention in (Yu, 2016; Furche, 2016; Hu and Vasilakos, 2016). Despite the achievement in technology or analysis process the problematic is still present. Indeed, we cannot neglect the great efforts and achievements made to improve the power of storage tools or methods of analysis to cover client expectations. But, getting the best information in record time remain a big challenge, knowing, this huge increase in data volume will have a critical impact on the overhead costs of computation, storage and networks as mention in some researches like (Kim *et al.*, 2017).

Based on different researches like (Rajasekar *et al.*, 2015) the number of data collected is continuously increasing, with sometimes the main goal is just to keep data under hand for an unknowing request in future. The way used for storage is still under argue by different parties as companies, politicians, people in power or ecologist.

Satisfying customer request is a normal behavior since the customer is king (Harris, 1991). However, to apply it remain an enormous task needs high tools and methods to target, their exact need and then extract the needful (Chen *et al.*, 2014a). For this reason, in our

approach we give high importance on the use of best practices to fit client's requests.

In this study we demonstrate step by step our approach used to conduct a Data Collect methodology despite domains. We first explain the need to focus on the collect data process despite the other processes present under data lifecycle in "Related works" section. For a better resulting we study in "Research Method" section the data lifecycle in literature focusing on data collect phase in each study. The "Result and Discussion" section demonstrate the equivalence between Systematic Literature Review (SLR) and Data Collect. Furthermore, the key steps of Systematic Literature Review were used as reference to conduct the phases of our proposed Data Collect Method with an example based on ODK suite. In the last section we will develop a conclusion of this work.

## Related Works

No one can argue about the high importance that information has in our life for all the fields. In a world where success depends on the quality, completeness and availability of the data on time for a purpose and not only having it with no benefit. The valuable information is the one responding to a specific need which will lead to the right decisions on time.

We principally build our data lifecycle based on

article (El Arass *et al.*, 2017) using the 11 phases, which are: Planning, Creation/Reception, Integration, Filtering, Anonymity, Enrichment, Analysis, Visualization, Storage, Destruction and Archiving. We consider the data Collect, all phases before Analysis Data. Hence, our strategy aims to improve the quality of all those phases coming before Analysis Data.

Therefore, to better understand the subject, Fig. 1 provide a global view of the lifecycle (Tikito and Souissi, 2017).

A number of researches focus on data lifecycle improvement to have the best quick solution for client request (El Arass *et al.*, 2017; El Arass and Souissi, 2018). Tracking this goal, several studies like [(Sangroya and Singhal, 2015; Cai *et al.*, 2016; Chen *et al.*, 2014b; Trnka, 2014; Cavanillas *et al.*, 2016) focus essentially on equipment for storage, computation and algorithms for the analysis to increase the capacity and speed. Knowing the common idea in different areas is, saving in case of we will need it sometimes, can lead to redundant data in different devices with different versions. Then analysing the huge amount of those data in all versions can in some cases significantly reduce the performance.

Furthermore, since many years several studies alert the saturation storage platforms despite the large number of IT equipment (Hilbert and Lopez, 2011; Ellis *et al.*, 2016) and until today researchers find a way to add more time,

then everyone continue saving data in different ways. Nonetheless, the emergence of cloud platforms, which is essentially a relocation of IT infrastructure, only hides the magnitude of the situation. Among the consequences of improper storage, a large number of servers are mobilized which impact the environment. In this context, the concept of "green computing" or sustainable computing was raised, which aims to reduce the secondary effects caused by computers (Gai *et al.*, 2016; Bisoyi and Das, 2016).

If we focus on gathering data without any control we can be sure our best processes today to analyse and manage data will fail on their commitment afterwards and a need to improve more all processes will be required. Our approach aims to collect less data but relevant, will straightforward use less storage space and increase data analysis and management performance. If we reduce the amount of data collected we can manage them quicker and better, making sure the efficiency and high quality of data. With guarantee of non-selected one don't fit the requirements.

Thus the reason why we propose a new method of data collection aim to make the process of collecting data "smart", consequently solve the problems encountered during data processing and the difficulties of storage and extraction. We will formalize the criteria to measure the intelligence of the data by presenting solutions to make it smarter before even being analysed.

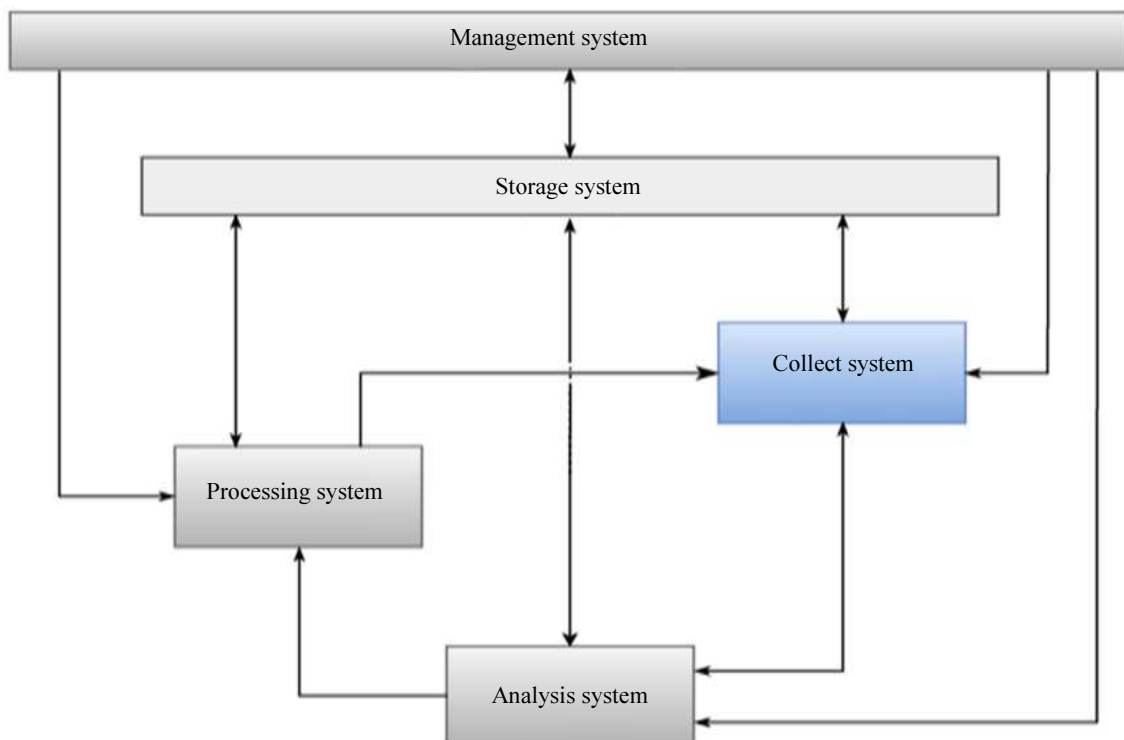


Fig. 1: Data lifecycle

The phase “Data Collect” is present in several articles only as an input for next phases without highlighting his importance, where our interest to improve it further. As mention in (Mcafee *et al.*, 2012) “You can’t manage what you don’t measure”. Our approach of making the collect data smart, help to reduce the volume of data by limiting unnecessary data to be extracted.

Data lifecycle present different steps that should be well-studied and optimized, to have a smart data. Many tools are implemented to analyze and integrate the most valuable data like (Wang *et al.*, 2016), respecting some criteria and answering to some purposes to be able to process and manage Big Data. While searching regarding the best way to optimise data lifecycle we notice the objective of gathering data has a similitude definition with Systematic Literature Review (SLR) method to collect all the data available respecting some predefined criteria to give an answer for a specific question.

The equivalence of “Data Collect” and “SLR” drive us to search further for the history and steps of this last finding. Light and Smith (1971) were the first to propose bringing together all relevant original data from various research studies. Thus, thanks to Archie Cochrane who has attracted the attention of researchers on the SLR method and those since 1972 when he published his famous book (Cochrane, 1972) emphasizing the quality of published research and the importance to have an overall view before making decisions.

However, the quality of an SLR is questioned in many cases due to misconceptions that are still persist and increase for the last years. Based on this situation happened for the SLR, we conduct a sufficient methodology to minimize the misconceptions due to a lack of explicit general methodology and based on a known concept. One of the criticisms that can be razed at researchers is that they define the SLR method in a specific field. We, therefore, In this study, relies on the methodology of SLR present in literature to conduct Data Collect with no domain condition in order to gain the greatest advantages from the process to establish our data collect process based on it.

## Research Method

In order to establish a method for a large number of audiences, we started with a literature study. (Colleen, 2015) in his book emphasizes the importance of the value of innovative approaches to the data collect. Thus, by performing analyzes early in the process the result will be more valuable.

Several publications define the data collection

method as a way to receive data as (Kothari, 2004), (Stembert *et al.*, 2013). Also (Alison and Susan, 2005) allows preparing questionnaires to collect data while being specific to a particular area, thus not allowing mutualisation. Similar to (Wilson *et al.*, 2016) they define the method followed for data collection in the field of in-depth interview for a very specific method:

- The researcher selecting the research participants;
- The researcher and participants co-scheduling the pre-paired depth interview sessions
- Participant-researchers co-constructing research questions during the pre-paired depth interview sessions
- Participant-researchers co-constructing the paired depth interview question(s) during the pre-paired depth interview sessions
- Participant-researchers conducting the first paired depth interview session involving each dyad being paired-interviewed by a selected participant-researcher
- Participant-researchers transcribing and distributing the transcripts from the first round of paired depth interview sessions
- The selected participant-researcher interviewers conducting the second paired depth interview session with the same dyad group to member-check the transcripts
- Each three-member paired-interview team of participant-researchers (i.e., interviewer and two interviewees) conducting the third paired depth interview session to co-analyze the paired depth interview data generated from the previous two sessions; and
- Participant-researchers deciding how they will document and disseminate the paired depth interview findings and interpretations

We illustrate in Table 1 a list of articles giving an importance to data collect phase in literature and how they describe it. Collect data definition turn around four basic notions:

- **Information source:** The articles define the phase as an input for another phase and not taking any interest on how these data are gathered or from where
- **Technology:** Cover all devices like connected objects designed to collect data
- **Methods:** Consist on approaches made for having data using for example a survey, Observations
- **Impact:** Evoke the power of data to change the decision

**Table 1:** Data Collect process definition around the literature

|                                   | Information Source | Technology | Methods | Impact |
|-----------------------------------|--------------------|------------|---------|--------|
| (U.S Brevet n° 6,078,924, 2000)   | X                  |            | X       |        |
| Clarke and Cossette (2000)        | X                  |            |         | X      |
| Kanari and Millar (2004)          |                    |            | X       |        |
| Witten and Eibe (2005)            | X                  |            |         |        |
| Mack (2005)                       | X                  |            | X       |        |
| (U.S. Brevet n° 7,086,603, 2006)  |                    | X          |         |        |
| (U.S. Brevet n° 7,406,436, 2008)  |                    | X          |         | X      |
| El Golli (2008)                   |                    | X          |         |        |
| (U.S. Brevet n° 7,551,922, 2009)  |                    | X          |         |        |
| Mertens (2009)                    | X                  |            | X       |        |
| Yu and Wen (2010)                 | X                  |            |         |        |
| Han <i>et al.</i> (2011)          | X                  | X          |         |        |
| (U.S. Brevet n° 8,036,929, 2011)  |                    | X          |         |        |
| Allard (2012)                     | X                  |            | X       |        |
| IBM (2013)                        | X                  |            |         |        |
| Peggy <i>et al.</i> (2014)        | X                  | X          |         |        |
| Creswell (2014)                   | X                  |            |         | X      |
| Lin <i>et al.</i> (2014)          | X                  | X          |         |        |
| Khan <i>et al.</i> (2014)         | X                  | X          | X       |        |
| Cigref (2014)                     | X                  | X          | X       |        |
| Demchenko <i>et al.</i> (2014)    | X                  | X          | X       |        |
| Fox <i>et al.</i> (2014)          | X                  |            |         |        |
| Faundeen <i>et al.</i> (2014)     | X                  | X          | X       |        |
| Khoufi <i>et al.</i> (2015)       |                    | X          |         | X      |
| Breckenridge <i>et al.</i> (2015) | X                  |            | X       |        |
| Benchimol <i>et al.</i> (2015)    | X                  |            | X       | X      |
| Michota and Katsikas (2015)       | X                  | X          | X       |        |
| Chaki (2015)                      | X                  |            |         |        |
| Siahaan <i>et al.</i> (2016)      | X                  |            | X       |        |
| Chaudhuri <i>et al.</i> (2016)    | X                  |            | X       |        |
| Cochrane (2016)                   | X                  | X          |         |        |
| Kalu (2017)                       |                    |            | X       | X      |
| Edmonds and Kennedy (2017)        | X                  |            | X       |        |

## Results

We were able to identify only two articles presenting a method of data collect: (Basili and Weiss, 1984) established five steps to follow:

- Establish the Goals of the Data Collection
- Develop List of Questions of Interest
- Establish Data Categories
- Design and Test Data Collection Form
- Collect and Validate Data

In last year only, another method not very different than the first one was defined by (Zozus, 2017):

- Identifying data to be collected
- Defining data elements
- Observing or measuring values or acquiring data by other means as in case of secondary data use
- Recording those observations and measurements

Processing data to render them in electronic form if not in electronic format already and prepare them for analysis.

The two methods have some similitude, but also are very close on the method of Systematic Literature Review. SLR is defined in (Cruz-Benito, 2016) as a type of literature review that collects and critically analyzes multiple research studies or documents through a systematic process. The equivalence of the two concepts can be resumed on the Table 2 based on followed notions: Context, Objective, Data source, Challenge and Importance. The notions used for mapping are selected to summarize the high-level definition about the two notions:

- **Context:** Define the circumstance of the notion
- **Objective:** The main goal of the notion
- **Data source:** The sources required or needed to collect the data needed
- **Importance:** The value of the notion mentions in literature

### *SLR Method: The Key Steps*

In this section, we summarize the different phases of SLR method described in (Tikito and Souissi, 2019):

- I. Pre-Planning
  1. A training relative to SR
  2. Identify the need of a review
  3. Specific question(s)
  4. Protocol
  5. Research strategy
  6. Data management
  7. Data collect
  8. Evaluation
  9. Commissioning
- II. Protocol
  1. Inclusion/exclusion criterion
    - a. Identify keywords
    - b. Linguistic constraints
    - c. Sources inclusion/exclusion
  2. Data evaluation criteria
  3. Integration criteria
  4. Requirements
  5. Methods to use
  6. Procedures to use
  7. Selection strategy
  8. Extraction strategy
- III. Data management
  1. Data collect
    - a. Selection studies
    - b. Data summary
    - c. Improve the collect
    - d. Managing references
    - e. Conclusion
  2. Subgroup management
    - a. Division criterion
    - b. Identify keywords
    - c. Subgroup analyzes
    - d. Interpretation
  3. Data analysis
    - a. Sensitivity analysis
    - b. Statistical analysis
    - c. Analysis data to exclude
    - d. Analysis contradictory research
    - e. Data comparison
    - f. Make a decision
    - g. Integration
  4. Evaluation
    - a. Evaluate quality criteria
    - b. Evaluate data quality
    - c. Evaluate integration criteria
  5. Synthesis
- IV. Present results
  1. Data description
    2. Deduce results
    3. Determine the gaps
    4. Data interpretation
    5. Summary
    6. Conclusion
    7. Recommendation
- V. Future research
  1. Define persisting problems
  2. Planning of future research
  3. Planning of future SLR
- VI. Publish
  1. Structure the report
  2. Write the report
  3. Evaluate the report
  4. Project calendar
  5. Specify publishing mechanisms
  6. Bibliography
  7. Appendix

### *Data Collect Method: The Key Steps*

Several publications define the data collection methodology as a way to receive data as (Kothari, 2004), (Stembert *et al.*, 2013). Also (Alison and Susan, 2005) allows preparing questionnaires to collect data while being specific to a particular area, thus not allowing mutualisation. We were able to identify only two articles presenting a methodology of data collection. Basili and Weiss (1984) established five steps to follow: establish the goals of the data collection; develop list of questions of interest; establish data categories; design and test data collection form; collect and validate data. In last year only, another methodology not very different than the first one was defined by (Zozus, 2017): identifying data to be collected; defining data elements; observing or measuring values, or acquiring data by other means as in case of secondary data use; recording those observations and measurements; processing data to render them in electronic form if not in electronic format already and prepare them for analysis.

Through the study of different methods proposed in the literature, we come out with a set of steps that can build a complete and detailed process for the creation of a data collect despite the field of research.

Resulting we did a semantic analysis to identify the meaning behind each step and dispelling any doubt. A codification of terms was requested to establish the steps for each phase.

**Table 2:** Mapping data collect Vs systematic literature review

|             | Data collect            | Systematic literature review             |
|-------------|-------------------------|--|
| Context     | Big Data                | Research studies or documents – Big Data |
| Objective   | Satisfy client request  | Answer research question                 |
| Data source | All                     | All                                      |
| Challenge   | Found the accurate data | Found the accurate data                  |
| Importance  | High                    | High                                     |

In order to establish a method responding to a general public, we have chosen to use the SLR method, since the latter has shown great efficiency and according to literature it aims to solve one or more problems by identifying, critically evaluating and integrating the results of all the studies (relevant and high-quality individual papers on one or more research questions). As a result, there is a similarity in the purpose of the SLR method and the collection of data. Accordingly, we establish the proposed method of Data Collect as follow based on the process present

Through the study of different methods proposed in the literature, we come out with a set of steps that can build a complete and detailed process for the creation of a data collect despite the field of research. Some names get change to answer the targeted context and to clarify any doubts as already explained.

The various steps are presented as follows:

- I. Pre-Planning
  1. Enquire Empathy Card
  2. Customer requirements
  3. Define Protocol strategy
  4. Define search strategy
  5. Define enrichment strategy
  6. Define storage strategy
  7. Define evaluation strategy
  8. Define Validation strategy
- II. Protocol
  1. Criteria to Include/Exclude
    - a. Define keywords
    - b. Sources to include/exclude
  2. Criteria for Integration
  3. Requirements based on 7V
  4. Methods to use
  5. Procedures to use
  6. Criteria for Satisfaction
- III. Data management
  1. Data Reception
    - a. Study selection
    - b. Summary of data
    - c. Improve collection
    - d. Manage sources
    - e. Smart data L0
  2. Data Creation
    - a. Criteria for creation
    - b. Define keywords
    - c. Data validation
    - d. Smart data L0
  3. Data Integration
    - a. Sensitivity analysis
    - b. Statistical analysis
    - c. Analyze negative result
    - d. Analyze contradictory result
    - e. Compare data
    - f. Structure data
    - g. Smart data L1
4. Evaluation
  - a. Evaluation of the Quality of the Criteria (EQC)
  - b. Evaluation of Data Quality (EDQ)
  - c. Evaluation of Integration Criteria (EIC)
5. Synthesis
- IV. Presentation
  1. Summary
  2. Deduce the results
    - a. Determine the gaps
  3. Interpret the data
  4. Define Persistent Problems (DPP)
  5. Define Improvement
- V. Enrichment
  1. Analyze the Need for Enrichment (ANE)
  2. Analyze the Need for Future Enrichment (ANFE)
- VI. Visualization
  1. Structure the deliverables
  2. Write the deliverables
  3. Report evaluation
  4. Publishing mechanisms
  5. Contract
  6. Process Improvements

Tests have been started with ODK 2 software suite by customizing the existing workflow with our proposed collect method. Based on the example below.

Pre-planning: In pre-planning we need to have a clear vision of client request and for our example we will take the case of a company wishing to collect data from a satisfaction survey.

Protocol: using ODK, we can define form template within criteria. Since GPS location is not needed for the survey, it will not be collected as information. Figure 2 shows the survey form.

Data management: many requests are implemented to handle different cases. Figure 3 displays the pop in when a user didn't fill all mandatory fields.

Or in case user is mention in our database with some criteria we will suggest him to use our smart collect instead of default collect, as shown in Fig. 4.

Presentation: ODK Aggregate allows presenting data in different ways.

Visualization: Using ODK Aggregate results we will implement a new form more efficient for users, answering client request. Figure 5 shows an example of Smart Collect Form pages and Fig. 6 is the visualization of the data collected using ODK.

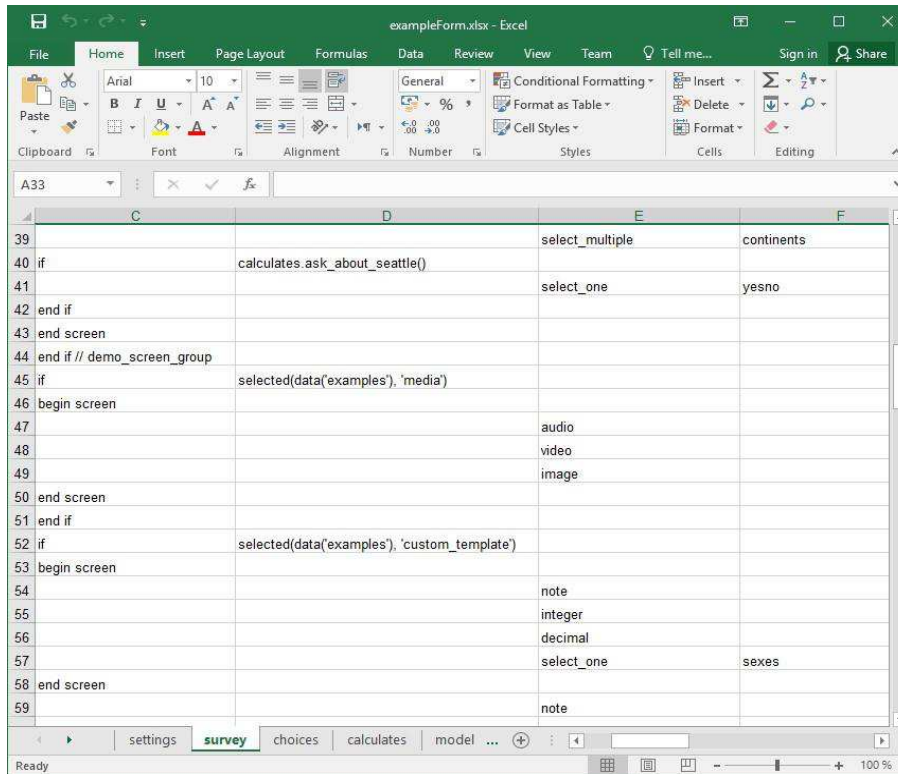


Fig. 2: ODK survey form

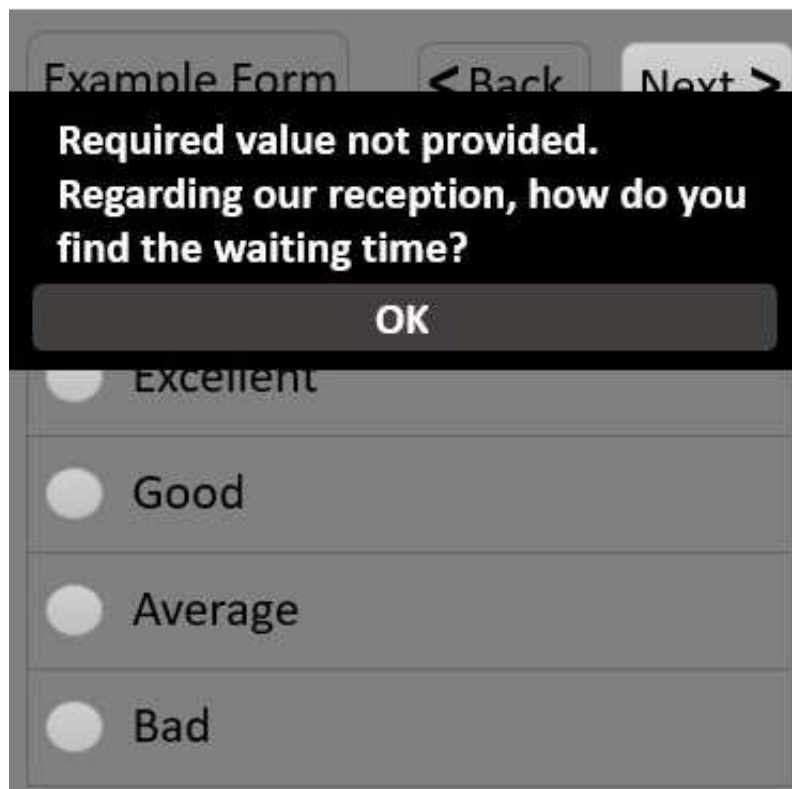


Fig. 3: Warning pop-in

### Choice of form

Do you want to start with “smart collect” or choose internal storage

INTERNAL STORAGE SMART COLLECT

Fig. 4: Choosing a form

Smart Collect Form < Back Next >

How did you find the kindness and listening of our agents?

|                                       |
|---------------------------------------|
| <input type="radio"/> Excellent       |
| <input checked="" type="radio"/> Good |
| <input type="radio"/> Average         |
| <input type="radio"/> Bad             |

Fig. 5: Smart Collect form





The screenshot shows the ODK Tables web interface. At the top, there are navigation tabs: Submissions, Form Management, ODK Tables (selected), and Site Admin. Below these are sub-tabs: Current Tables, View Table, Manage Instance Files, Manage Table Files, and Manage App Level Files. A dropdown menu shows 'exampleForm' selected. Below the dropdown, it says 'Displaying: exampleForm'. A table lists files with columns: Row ID, Filename, Size, Content Type, and Download. The table contains three rows of data.

| Row ID  | Filename | Size       | Content Type        | Download |
|---|----------|------------|---------------------|----------|
| uuid:c045ab85-127c-46fb-8249-6fed2075c682_1544952242684_android_id_3bb274448d7974d7.mp3 | 3283989  | audio/mpeg | <a href="#">Get</a> |          |
| uuid:c045ab85-127c-46fb-8249-6fed2075c682_1544952300263_android_id_3bb274448d7974d7.mp4 | 2067106  | video/mp4  | <a href="#">Get</a> |          |
| uuid:c045ab85-127c-46fb-8249-6fed2075c682_1544952317468_android_id_3bb274448d7974d7.jpg | 4703346  | image/jpeg | <a href="#">Get</a> |          |

Fig. 6: Presentation form

## Conclusion

With the progress of Big data era, improving the data lifecycle is becoming unavoidable choice for researches. The importance of data is well known in all fields even if some are more significant than other. In this study, we proposed a new Data Collect process to be the first step for improvement. We structured the phases based on Systematic Literature Review method due to the similitude between the two concepts. We suggest a definition of Big data after analyzing different researches and simplifying the terms to fit with all domains and levels.

In this study, we were able to spotlight the importance of Data Collect process in Data lifecycle and provide a first process based on literature review. To conduct this result, we established first a complete Systematic Literature Review process and then we conclude the Data Collect method.

Our future work consists to validate the efficient of this method in a real field toward evaluate the advantage on return. First step we will develop the flow based on ODK v2.0 software for collecting, managing and using data in resource-constrained environments. The gathered data from different resources will then endure the regular process after using our suggested process to compare the quality.

## Acknowledgement

We are very grateful to each person helps us to reach this result.

## Author's Contributions

**Iman Tikito and Nissrine Souissi:** Have equally contributed to this manuscript.

**Mohammed El Arass:** Participated in the analysis and contributed to the writing of the manuscript.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

## References

- Alison, M. and M.G. Susan, 2005. Second Language Research: Methodology and Design. 2nd Edn., Routledge, ISBN-10: 1317612655, pp: 434.
- Allard, S., 2012. DataONE: Facilitating eScience through collaboration. J. eScience Librarianship, 1: 3. DOI: 10.7191/jeslib.2012.1004
- Basili, V. R. and Weiss D. M., 1984. A methodology for collecting valid software engineering data. IEEE Trans. Software Eng., 6: 728-738.
- Benchamol, E.I., L. Smeeth and A. Guttman, 2015. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med., 12: e1001885. DOI: 10.1371/journal.pmed.1001885
- Bisoyi, B. and B. Das, 2016. Necessitate green environment for sustainable computing. Proceedings of the 2nd International Conference on Computer and Communication Technologies. Sept. 4-4, Springer India, pp: 515-524. DOI: 10.1007/978-81-322-2523-2\_50
- Breckenridge, K., H.L. Bekker and E. Gibbons, 2015. How to routinely collect data on patient-reported outcome and experience measures in renal registries in Europe: An expert consensus meeting. Nephrology Dialysis Transplantation, 30: 1605-1614. DOI: 10.1093/ndt/gfv209.
- Cai, H., B. Xu and L. Jiang, 2016. IoT-based big data storage systems in cloud computing: Perspectives and challenges. IEEE Int. Thi. J. 4: 75-87. DOI: 10.1109/JIOT.2016.2619369

- Cavanillas, J.M., E. Curry and W. Wahlster, 2016. New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe. 1st Edn., Springer, ISBN-10: 3319215698, pp: 303.
- Chaki, S., 2015. The lifecycle of enterprise information management. Enterprise Information Management Practice.  
DOI: 10.1007/978-1-4842-1218-9\_2
- Chaudhuri, A., T.C. Christofides and R. Rao Calyampudi, 2016. Data gathering, analysis and protection of privacy through randomized response techniques: Qualitative and quantitative human Traits. s.l.:Elsevier.
- Chen, M., S. Mao and Y. Liu, 2014a. Big data: A survey. Mobile Networks Applic., 19: 171-209.  
DOI: 10.1007/s11036-013-0489-0
- Chen, M., S. Mao and Y. Zhang, 2014b. Big Data Related Technologies, Challenges and Future Prospects. 1st Edn., Springer, ISBN-10: 978-3-319-06245-7, pp: 99.
- Cigref, 2014. Enjeux business des données. s.l.:s.n.
- Clarke, S.P. and S. Cossette, 2000. Secondary analysis: Theoretical, methodological and practical considerations. Canadian J. Nursing Res. Archive.
- Cochrane, A.L., 1972. Effectiveness and Efficiency: Random Reflections on Health Services. 1st Edn., Nuffield Provincial Hospitals Trust.
- Cochrane M., 2016. Turn your phones on: using Android devices to collect scientific data.
- Colleen, M., 2015. Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis. 2 Edn., Elsevier Science, ISBN-10: 0128002298, pp: 422.
- Creswell, J.W., 2014. Research Design: Qualitative, quantitative and mixed methods approaches. s.l.:Sage.
- Cruz-Benito, J., 2016. Systematic literature review and mapping. PhD Thesis, Department of Computers and Automatics University of Salamanca, Salamanca, Spain. DOI: 10.5281/zenodo.165773
- Demchenko, Y., De Laat, C. and Membrey, P., 2014. Defining architecture components of the big data ecosystem. Collaboration Technologies and Systems (CTS), pp: 104-112.  
DOI: 10.1109/CTS.2014.6867550.
- Edmonds, W.A. and T.D. Kennedy, 2017. An Applied Guide to Research Designs: Quantitative, Qualitative and Mixed Methods. s.l.:Sage.
- El Arass, M. and N. Souissi, 2018. Data Lifecycle: From Big Data to SmartData. Proceedings of the IEEE 5th International Congress on Information Science and Technology (CiSt), Marrakech, pp: 80-87.  
DOI: 10.1109/CIST.2018.8596547
- El Arass, M., I. Tikito and N. Souissi, 2017. Data lifecycles analysis: Towards intelligent cycle. Proceedings of the Intelligent Systems and Computer Vision, Apr. 17-19, IEEE Xplore Press, Fez, Morocco, pp: 119-125.  
DOI: 10.1109/ISACV.2017.8054938
- El Golli, I.G., 2008. Ingénierie des exigences pour les systèmes d'information décisionnels: Concepts, modèles et processus - la méthode CADWE, Paris 1: PhD Thesis.
- Ellis, A.D., N.M. Suibhne, D. Saad and D.N. Payne, 2016. Communication networks beyond the capacity crunch. Phil. Trans. R. Soc., 6: 374.  
DOI: 10.1098/rsta.2015.0191
- Faundeen, J.L., 2014. The united states geological survey science data lifecycle model: U.S. geological survey open-file report, pp: 2013-1265.
- Fox, MA., M.C. Sheehan, T.A. Burke, A. Navas-Acien and P.N. Breyse *et al.*, 2014. Global methylmercury exposure from seafood consumption and risk of developmental neurotoxicity: A systematic review. Bulletin World Health Organization, 92: 254-269F.
- Furche, T., 2016. Data Wrangling for big data: Challenges and opportunities. Proceedings of the 19th International Conference on Extending Database Technology, Mar. 15-15, Bordeaux, France, pp: 473-478.  
DOI: 10.5441/002/edbt.2016.44
- Gai K., Qiu M. and Zhao H., 2016. Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing. J. Network Comput. Applic., 59: 46-54.
- Han J., Pei J. and Kamber M., 2011. Data mining: concepts and techniques. s.l.:Elsevier.
- Harris, R.L., 1991. The Customer is King. 1st Edn., ASQC Quality Press, ISBN-10: 0873891228, pp: 144.
- Hilbert, M. and P. Lopez, 2011. The world's technological capacity to store, communicate and compute information. Science, 332: 60-65.
- Hu, J. and A. V. Vasilakos, 2016. Energy big data analytics and security: Challenges and opportunities. IEEE Trans. Smart Grid, 7: 2423-2436.
- IBM, 2013. Wrangling big data : Fundamentals of data lifecycle management, s.l.: s.n.
- Kalu, F. A., 2017. What makes qualitative research good research? An exploratory analysis of critical elements. Int. J. Social Sci. Res., 5: 43-56.  
DOI: 10.5296/ijssr.v5i2.10711
- Kanari, Z. and R. Millar, 2004. Reasoning from data: How students collect and interpret data in science investigations. J. Res. Sci. Teaching, 41: 748-769.  
DOI: 10.1002/tea.20020.
- Khan, N., I. Yaqoob, I.A.T. Hashem and Z. Inayat and M. Ali *et al.*, 2014. Big data: survey, technologies, opportunities and challenges. Scientific World J.

- Khoufi, I., S. Mahfoudh, P. Minet and A. Laouiti, 2015. Data gathering architecture for temporary worksites based on a uniform deployment of wireless sensors. *Int. J. Sensor Netw.*, 18: 3-21. DOI: 10.1504/IJSNET.2015.069874
- Kim, D., S. Song and B.Y. Choi, 2017. *Data Deduplication for Data Optimization for Storage and Network Systems*. 1st Edn., Springer, ISBN-10: 3319422804, pp: 262.
- Kothari, C.R., 2004. *Research Methodology: Methods and Techniques*. 1st Edn., New Age International, ISBN-10: 8122415229, pp: 401.
- Light, R. and P. Smith, 1971. Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Edu. Rev.*, 41: 429-471. DOI: 10.17763/haer.41.4.437714870334w144
- Lin, L., Liu, T., Hu, J. and Zhang, J., 2014. A privacy-aware cloud service selection method toward data life-cycle. *Proceedings of the 20th IEEE International Conference Parallel and Distributed Systems (ICPADS)*, pp: 752-759. DOI: 10.1109/PADSW.2014.7097878
- Mack, N., 2005. *Qualitative research methods: A data collectors field guide*. s.l.:Family Health International.
- Mcafee, A., E. Brynjolfsson and T.H. Davenport, 2012. Big data: The management revolution. *Harvard Bus. Rev.*, 90: 60-68.
- Mertens, D.M., 2009. *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative and mixed methods*. s.l.:Sage.
- Michota, A. and S. Katsikas, 2015. Designing a seamless privacy policy for social networks. *Proceedings of the 19th Panhellenic Conference on Informatics*, pp: 139-143. DOI: 10.1145/2801948.2801998
- Peggy, S., L. Vincent and V. Eric, 2014. Using event data recorder to detect road infrastructure failures from a safety point of view. *Proceedings of the European Transport Conference, Association for European Transport*.
- Rajasekar, D., C. Dhanamani and S.K. Sandhya, 2015. A survey on big data concepts and tools. *Int. J. Emerg. Technol. Adv. Eng.*, 5: 80-81.
- Sangroya, A. and R. Singhal, 2015. Performance assurance model for HiveQL on large data volume. *Proceedings of the 22nd International Conference on High Performance Computing Workshops*, Dec.16-19, IEEE Xplore Press, Bangalore, India, pp: 26-33. DOI: 10.1109/HiPCW.2015.8
- Siahaan, E., A. Hanjalic and J. Redi, 2016. A reliable methodology to collect ground truth data of image aesthetic appeal. *IEEE Trans. Multimedia*, 18: 1338-1350. DOI: 10.1109/TMM.2016.2559942
- Stembert, N., P. Conradie, I. Mulder and S. Choenni, 2013. Participatory data gathering for public sector reuse: Lessons learned from traditional initiatives. *Proceedings of the International Conference on Electronic Government (CEG' 13)*, Springer, Berlin, Heidelberg, pp: 87-98. DOI: 10.1007/978-3-642-40358-3\_8
- Tikito, I. and N. Souissi, 2017. Data collect requirements model. *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications*, Mar. 29-30, Tetouan, Morocco, ACM, p: 4-4. DOI: 10.1145/3090354.3090358
- Tikito, I. and N. Souissi, 2019. Meta-analysis of systematic literature review methods. *Int. J. Modern Edu. Comput. Sci.*, 11: 17-17. DOI: 10.5815/ijmecs.2019.02.03
- Trnka, A., 2014. Big data analysis. *Eur. J. Sci. Theol.*, 10: 143-148.
- Wang, Y., L. Kung and T.A. Byrd, 2016. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Change*, 126: 3-13. DOI: 10.1016/j.techfore.2015.12.019
- Wilson, A.D., A.J. Onwuegbuzie and L.P. Manning, 2016. Using paired depth interviews to collect qualitative data. *Qualitat. Report*, 21: 1549-1573.
- Witten, I.H. and F. Eibe, 2005. *Data Mining: Practical machine learning tools and techniques*. s.l.:Morgan Kaufmann.
- Yu, S., 2016. Big privacy: Challenges and opportunities of privacy study in the age of big data. *IEEE Access*, 4: 2751-2763.
- Yu, X. and Q. Wen, 2010. A view about cloud data security from data lifecycle. *Computational Intelligence and Software Engineering (CiSE)*, pp: 1-4. DOI: 10.1109/CISE.2010.5676895
- Zozus, M., 2017. *The Data Book: Collection and Management of Research Data*. 1st Edn., CRC Press, Canada, ISBN-10: 1351647733, pp: 336.