

Article

A Comprehensive Framework to Reinforce Evidence Synthesis Features in Cloud-Based Systematic Review Tools

Tatiana Person ^{1,*} , Iván Ruiz-Rube ¹ , José Miguel Mota ¹ , Manuel Jesús Cobo ¹ , Alexey Tselykh ²  and Juan Manuel Dodero ¹ 

¹ Department of Informatics Engineering, University of Cadiz, 11519 Puerto Real, Spain; ivan.ruiz@uca.es (I.R.-R.); josemiguel.mota@uca.es (J.M.M.); manuel.cobo@uca.es (M.J.C.); juanma.dodero@uca.es (J.M.D.)

² Department of Information and Analytical Security Systems, Institute of Computer Technologies and Information Security, Southern Federal University, 347922 Taganrog, Russia; tselykh@sfedu.ru

* Correspondence: tatiana.person@uca.es

Abstract: Systematic reviews are powerful methods used to determine the state-of-the-art in a given field from existing studies and literature. They are critical but time-consuming in research and decision making for various disciplines. When conducting a review, a large volume of data is usually generated from relevant studies. Computer-based tools are often used to manage such data and to support the systematic review process. This paper describes a comprehensive analysis to gather the required features of a systematic review tool, in order to support the complete evidence synthesis process. We propose a framework, elaborated by consulting experts in different knowledge areas, to evaluate significant features and thus reinforce existing tool capabilities. The framework will be used to enhance the currently available functionality of CloudSERA, a cloud-based systematic review tool focused on Computer Science, to implement evidence-based systematic review processes in other disciplines.

Keywords: systematic review tool; systematic review protocol; evidence synthesis methods; evidence-based research



Citation: Person, T.; Ruiz-Rube, I.; Mota, J.M.; Cobo, M.J.; Tselykh, A.; Dodero, J.M. A Comprehensive Framework to Reinforce Evidence Synthesis Features in Cloud-Based Systematic Review Tools. *Appl. Sci.* **2021**, *11*, 5527. <https://doi.org/10.3390/app11125527>

Academic Editor: Miguel Ángel Conde

Received: 27 May 2021
Accepted: 11 June 2021
Published: 15 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Research and development activity and decision making usually require a preliminary study of related literature to understand the up-to-date, state-of-the-art issues, techniques and methods in a given research field. For instance, Health Science researchers need to find out the scientific evidence that supports their clinical decisions. Analyses in bibliometrics [1], science mapping [2,3] and logology [4] need to operate on data records that are usually retrieved from queries to a bibliographic database, such as Clarivate's Web of Science or Elsevier's SCOPUS, or a patent registry. A huge volume of data is published and stored in digital bibliographic repositories, which are often manually reviewed in order to select those related to the field and research purpose. Thus, it is important to be acquainted with the quality of the evidence provided in these studies. In this vein, tools such as GRADEpro [5] have emerged, to synthesize and evaluate the quality of evidence found in health science-related studies.

Rooted in the Health Sciences, Evidence Synthesis (ES) methods are used to aggregate the global message of a set of studies [6]. The main goal of ES is to evaluate the included studies and select appropriate methods for integrating their information [7]. ES methods can be used to synthesize both qualitative and quantitative evidence [8], according to the type of research questions and forms of evidence analyzed. These methods are often specific or adapted to a given field. For example, scoping, thematic analysis, narrative synthesis, comparative analysis, meta-analysis, case survey and meta-ethnography are ES methods in the Software Engineering field [9].

Evidence synthesis approaches are seamlessly linked to Systematic Reviews (SR) methods, which enable researchers to identify, evaluate and interpret the existing research that is relevant for a particular Research Question (RQ) or phenomenon of interest [10]. Some reasons for performing SRs include: to synthesize the existing evidence concerning a given topic, to identify gaps in the current research and to suggest areas for further investigation; and to provide a background for positioning new research lines [11].

Focused on the disciplines of the Health Sciences, the ES methods' steps are defined as the following [12]: aggregate information; explain or interpret processes, perceptions, beliefs and values; develop theory; identify gaps in the literature or the need for future research; explore methodological aspects of a method or topic; and develop or describe frameworks, guidelines, models, measures, scales or programmes. SR methods have also been used in domains such as Environmental Sciences [13] and Computer Science [14], which have benefited from the ES approach. In the latter field, a set of guidelines for performing Systematic Literature Reviews (SLR) has been published [14]. The guidelines define an SLR as a process consisting of three stages, namely, planning, conducting and reporting. The SLR method has become a popular research methodology for conducting literature reviews and evidence aggregation in Software Engineering. Similarly, Systematic Mapping Studies (SMS) and scope studies enable researchers to obtain a wide overview of a research area, providing them with a quantitative indication of the evidence found [15].

It is important to note that, regardless of the discipline in which ES is applied, a considerable number of studies must be processed and, eventually, selected as primary. Consequently, the information provided for those studies should be methodically synthesized. This process is a time-consuming task and is difficult to conduct manually. For this reason, using computer tools to support the process is essential in research and decision-making.

The goal of this paper is to analyse and collect the essential features of ES methods in order to present a framework that can be used to improve cloud-based SR support tools, thus fostering comprehensive evidence-based systematic review processes. The main contribution is a framework that aggregates cloud-based ES features as proposed and used in existing SR tools. To accomplish this goal, a design and creation research strategy [16] has been followed and applied around the CloudSERA software artifact. CloudSERA [17] is a cloud-based web application that supports systematic reviews of scientific literature. Its current version is focused on SLR processes applied in Computer Science. The tool has been previously evaluated in the Computer Science discipline under the scope of the SLR methods. For the sake of generality, the features of future versions of CloudSERA have to be proposed and assessed within other research domains beyond Computer Science.

The research output of the design and creation strategy is a framework or construct that covers the concepts and vocabulary [16] used in the ES and SR domains. This construct is the basis of instantiations or working systems, such as the future version of CloudSERA. As is common in computing and information systems research, the methodology involves analysing, designing and developing a computer-based product to explore and exhibit the possibilities of software technologies applied to the SR domain. This work does not approach an illustration of technical prowess in the development of the software artifacts, but instead an analysis, argument and critical evaluation [16] of the features to be considered for augmenting CloudSERA. Therefore, a survey with experts from Computer Science and other disciplines has been performed to collect their opinions about the features included in the proposed framework, and to gather new necessary features with the objective of applying them in the next versions of CloudSERA, reinforcing its functionality to enable the ES process and using it in a multidisciplinary way.

2. Evidence-Based Systematic Review Analysis Framework

Following the aim to define a framework that provides the set of features needed in a tool to cover an evidence-based SR process, we analyzed the existing tools that support such processes. First, we have based our framework on the prior analysis of Kohl [18], who carried out an evaluation of existing SR tools according to a set of features selected

from previous studies. Second, we considered the work of Hassler [19], who conducted a community workshop with Software Engineering researchers in order to identify and prioritize the necessary features of an SLR tool. Finally, the work of Manterola [20] provided us with a suite of ES features as the relevant steps to be carried out in ES methods. Based on such previous studies, we defined an analysis framework, as shown in Figure 1, using the BPMN notation. It is important to note that two participant profiles have been defined in the framework, namely researcher and data scientist, because of the complexity of the complete process. The same individuals can perform both roles as long as they have learned specific data analysis techniques and tools. For instance, Cheng [21] notes the importance of using machine learning techniques for applying ES methods in conservation and environmental studies. Based on the proposed framework, each main feature required for an integrated evidence-based SR tool is classified in one of the following categories (see Table 1):

- **Non-functional:** This category contains non-functional features related to the SR tool, such as open source code availability, licensing mode (and cost), availability of user guides, focus on specific disciplines, etc.
- **Overall functionality:** This category contains features that, not being specific to SR tools, can be interesting and make the task easier. These include the capability for collaborative work, management of user roles, maintenance of data history and traceability of the SR process steps.
- **Information management:** This category includes SR-specific features related to the steps for studies' inclusion, selection and management. These steps include: definition of the research questions and scope of the study; definition and running of queries; study selection; and application of quality assessment criteria over the studies.
- **Evidence synthesis:** This category includes features related to the synthesis of relevant data and information obtained from the selected studies. The steps required for this are the following [20]: choosing the type of synthesis (either qualitative or quantitative); selecting the synthesis method; defining the study variables to measure; tagging and extraction of information from studies; critical appraisal of the risk of biased assessments; configuring the individual values from selected studies to calculate the global values of variables to measure; and generating reports with the synthesized data gathered.

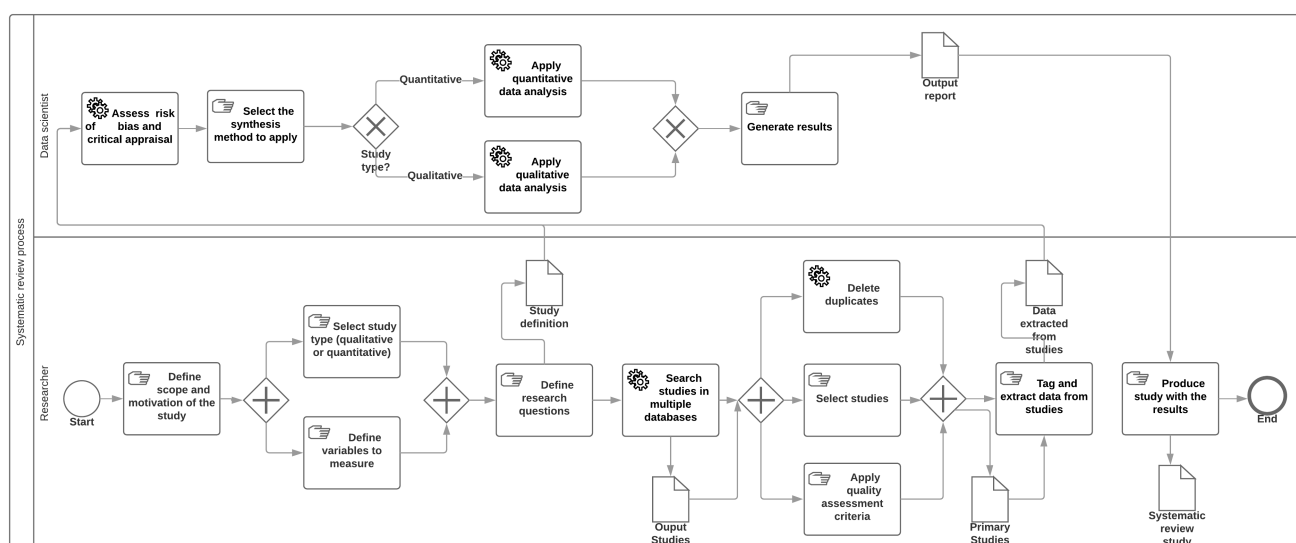


Figure 1. Systematic review process framework.

Table 1. Classification of main needed features in SR support tools.

Feature	Description
Non-functional	
F1. Cloud	Accessible for use online in web
F2. Open source & Free	Availability of the source code. Free of charge
F3. Updated	Tool maintenance carried out less than a year ago
F4. Not focused	Not focused on any particular academic discipline
F5. User guides	User and installation guides, tutorials, and any type of resources that guide users to use the tool easily
Overall functionality	
F6. Collaboration	Collaboration to perform the SR tasks with other users
F7. User role management	Creating and management of user roles and permissions in an SR project
F8. Data maintenance	Data maintenance and preservation functions to access past research questions, protocols, studies, data, metadata, bibliographic data and reports
F9. Traceability	Forward and backward traceability to link goals, actions, change history and results for accountability, standardization, verification and validation
Information management	
F10. Research question	Ability to define research questions and related problems
F11. Scope	Ability to define the scope of the study
F12. Integrated search	Ability to search multiple databases without having to perform separate searches
F13. Duplicate deletion	Automatic deletion of duplicate studies
F14. Study selection	Selection of primary studies using inclusion/exclusion criteria
F15. Quality assessment	Evaluation of primary studies using quality assessment criteria
Evidence synthesis	
F16. Study type	Definition of the type of the study based on the qualitative or quantitative analysis to apply
F17. Synthesis method	Selection of the synthesis method to apply (according to the study type selected)
F18. Study variables	Definition of the variables to measure and their types
F19. Tag and data extraction	For quantitative studies extraction of data and tags automatically from studies
F20. Risk of bias and critical appraisal	Assess risk of bias and critical appraisal in individual studies
F21. Summarizing	Summarizing results from studies data
F22. Reporting	Adapted for the analysis type (graph type and statistical results more adequate in each case)

Table 2 show the results of an analysis made on a set of SR tools that are not focused on specific research fields [18]. Thus, the selected tools do not depend on the particular features required in a specific discipline. These tools are: CADIMA [18], Colandr [22], DistillerSR [23], EPI-Reviewer [24], METAGEAR R package [25], Rayyan [26], ReviewER [27], SESRA [28], SLR-Tool [29] and CloudSERA [17]. Subsequently, the results obtained are discussed for each tool and feature category.

Table 2. Analysis of main desirable features in existing SR tools.

	Non-Functional				Overall Functionality					Information Management							Evidence Synthesis					
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22
CADIMA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓		✓	✓	✓
Colandr	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓				✓		✓	✓	✓
DistillerSR	✓		✓	✓	✓	✓		✓					✓	✓	✓			✓		✓	✓	✓
EPI-Reviewer			✓	✓	✓	✓		✓	✓				✓	✓				✓		✓	✓	✓
METAGEAR R		✓	✓	✓	✓	✓		✓										✓			✓	✓
Rayyan	✓	✓	✓	✓	✓	✓	✓					✓	✓									
ReviewerR		✓		✓		✓		✓					✓	✓								
SESRA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓				✓			✓	✓
SLR-Tool		✓		✓				✓	✓	✓	✓		✓	✓				✓		✓	✓	✓
CloudSERA	✓	✓	✓		✓	✓		✓	✓	✓	✓	✓	✓	✓	✓							✓

From the feature analysis shown in Table 3, we can see that some tools are more focused on the management of information related to selected studies, such as ReviewER or CloudSERA. Other tools are focused on information synthesis, such as the METAGEAR R package. Additionally, CADIMA, SLR-Tool and Colandr are balanced in both categories. We can notice how there is no tool that fully covers all requirement categories. In addition, the evidence synthesis category has an average coverage score of 38.54% and the maximum obtained by a tool is 57.1%. This might indicate that more work needs to be done on re-searching and improving these types of features in the SR process support tools. Eventually, the result of the analysis on the current version of CloudSERA yields the following scores:

- Non-functional features: 80%.
- Overall functionality: 75%.
- Information management: 100%.
- Evidence synthesis: 14.28%.
- Average score: 67.32%.

Table 3. Percentage results of the comparison of available features in SR process support tools.

Tool	Non-Functional	Overall Functionality	Information Management	Evidence Synthesis
CADIMA	100	100	83.3	57.1
Colandr	100	100	50	57.1
DistillerSR	80	50	50	57.1
EPI-Reviewer 4	60	75	33.3	57.1
METAGEAR R	80	50	0	42.8
Rayyan	100	75	33.3	0
ReviewER	40	50	33.3	0
SESRA	100	100	50	42.8
SLR-Tool	40	50	66.6	57.1
CloudSERA	80	75	100	14.28
Average	78	72.5	49.98	38.54

The previous results would guide us towards the new functionalities that must be incorporated in CloudSERA, besides identifying those that are essential for covering the complete ES process, which is the main goal of this work.

3. Cloudsera Features and Implementation

From the previous results, none of the analyzed SR tools cover the complete SR process. For this reason, we plan to expand CloudSERA to incorporate the complete set of evidence synthesis features. The current state of CloudSERA functionality is described below.

3.1. Cloudsera Features

Since it is a cloud-based web application, CloudSERA does not require installation or configuration. It is available online for free usage [30]. Besides, CloudSERA is an open-source tool, with its source code openly released in GitHub [31].

Concerning the features of the non-functional category, the Grails framework has been used to develop the application. Figure 2 shows the conceptual information model managed by the tool. The user interface has been built using the Bootstrap toolkit, which provides a responsive and rich user experience. The tool is provided with development documentation and end-user tutorials. To summarize, the tool has been developed covering the complete set of non-function features, namely, it is cloud-based, open source and free, updated, not focused on a specific discipline, and delivered with user guides.

Considering the overall category features, the tool has been implemented with a role management module, thus enabling users to collaboratively work on a review. Two main roles are defined, namely performer and supervisor. SR data can be shared among all the SR team members. With a user's consent, the SR data can be accessed through the

web interface for preservation and reproducibility (see Figure 3b). Thus, users can follow other users' activities. Finally, CloudSERA provides a logging system to trace the actions performed by users in an SR project. To summarize, CloudSERA has been thought of, and built, to cover the main features of the overall functionality category, such as collaboration, user role management, data maintenance and traceability.

Regarding the information management category features, users can create SRs and define research questions and related issues. CloudSERA can be used to automate several tasks of the SR process and includes a step-wise wizard (see Figure 3a) to guide users through the creation and configuration of an SR process. Besides, the tool automates the search tasks by launching the configured queries to an integrated set of databases and digital libraries. Currently, the supported sources are the following: ACM Digital Library, IEEE Computer Society, Springer Link and Science Direct. CloudSERA enables the inclusion of new sources easily through configuration. With the integrated search engines, the user does not have to run separate queries for each library. Every query runs asynchronously in the background and the user is notified when the search finishes. CloudSERA enables the definition of inclusion and exclusion criteria, which can be used to filter the bibliographic references found. References can be visualized and refined by means of a set of facets, according to the automatically retrieved metadata of the studies and the manually entered values for the attributes (see Figure 3c). These results will also be used to show the statistics of included and excluded studies, with their exclusion reasons.

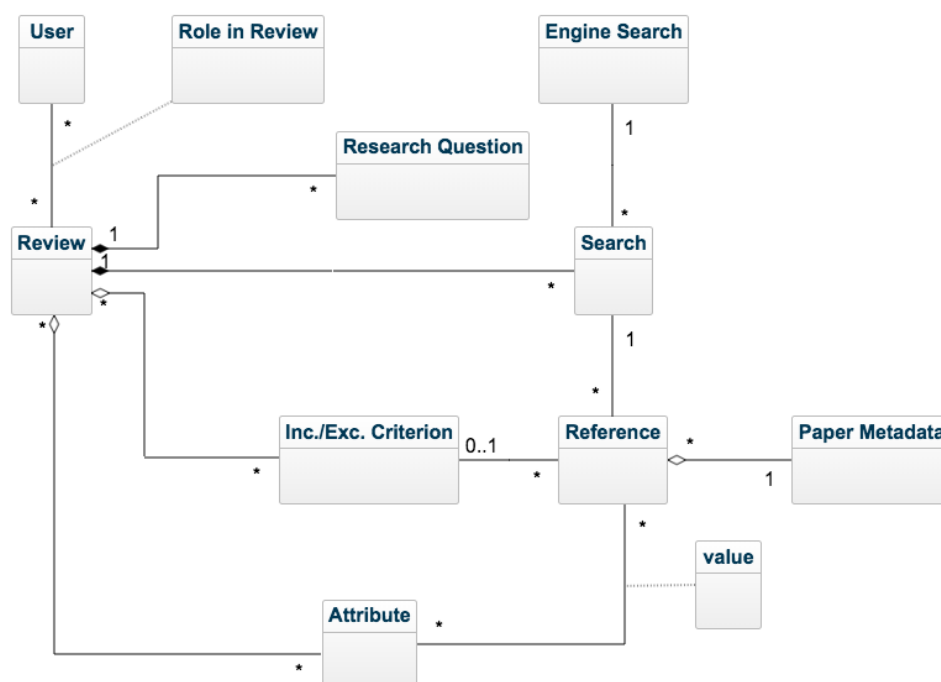


Figure 2. CloudSERA conceptual model.

Considering the features of the evidence synthesis category, CloudSERA uses charts to visualize the data according to some aspects such as document type, language and inclusion or exclusion criteria, among others. Figure 3d shows an example of the main screen of a specific project. This enables users to report data results from the study inclusion/exclusion criteria and specific attribute tagging.

Finally, Elsevier's Mendeley is also integrated with CloudSERA to authenticate users using Mendeley credentials and to import and store bibliographic references found. Common metadata is used to automatically annotate the imported references. The tool also adds specific attributes for designing additional data extraction forms and quality assessment instruments. In this way, users can collect all the information needed from the primary

studies to address the review questions by using textual or nominal attributes in a range of predefined values. In addition, users can evaluate the quality of each compiled study by means of a numeric attribute-based scale. These attributes will be used to categorize the studies and export the results, including the statistics of the studies tagged by each category. Additionally, the tool enables the user to export the bibliography data in different formats, for example, BibTeX, Word and Excel. The two latter formats also provide pages or sheets that include the resulting data of the work, such as research questions, attributes, search history, primary studies and charts (see the available export options in Figure 3d).

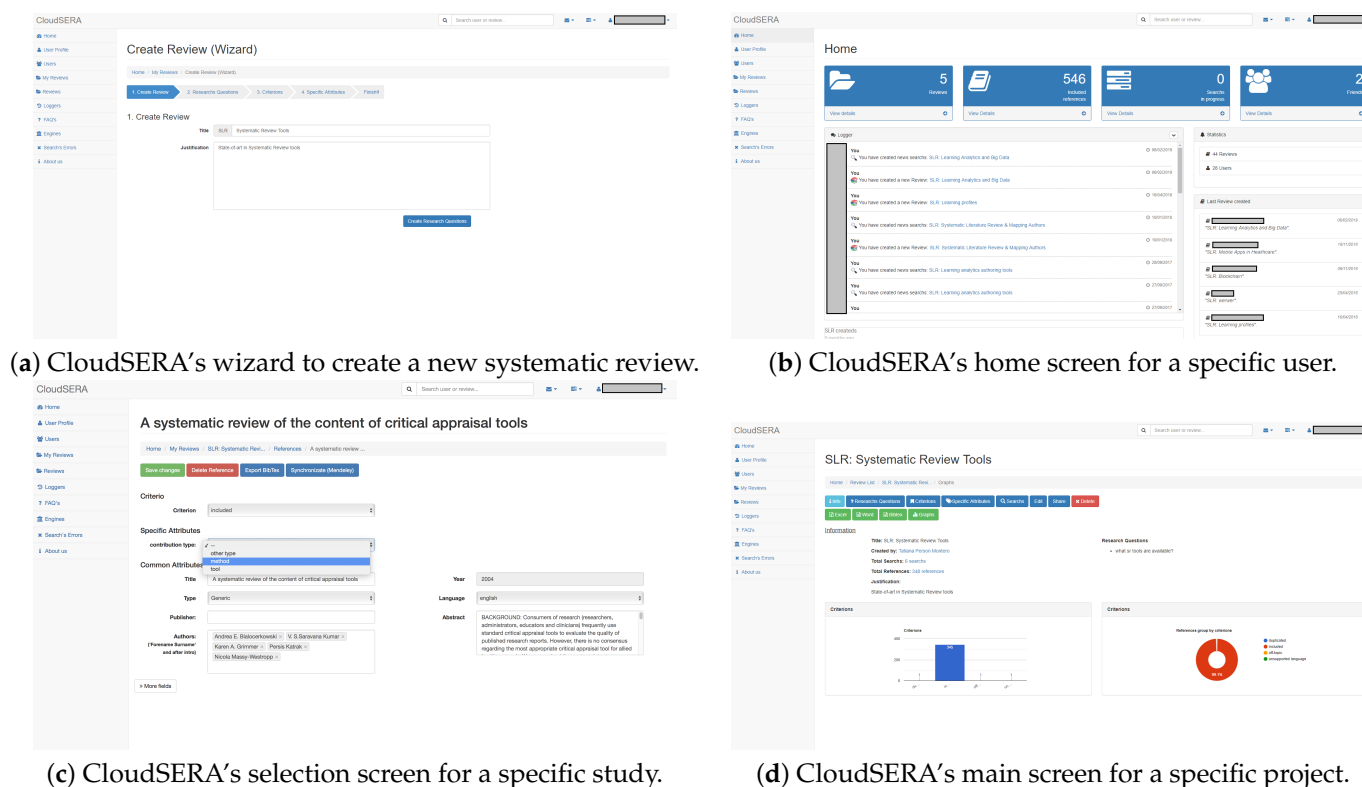


Figure 3. CloudSERA's main screens.

3.2. Technical Quality and Utility Evaluation

The technical quality and utility of software must be rigorously checked using accepted evaluation methods [32]. CloudSERA has been evaluated, first, by means of software quality testing techniques. With that aim, an exhaustive test battery was developed and run to ensure the fulfillment of the requirements. Sets of unit tests and functional tests were coded by using Spock and Selenium frameworks. Then, several non-functional tests were also conducted. JMeter was used first to stress-test the system and check its behavior with a great number of requests and long reference searches; then TAWDIS was used to check the web accessibility. Finally, a structural inspection of code quality to detect bugs, code smells and security vulnerabilities was performed with SonarQube. More details about the testing plan can be found in the developer portal of the tool [31].

Once the application was developed and deployed, a general heuristic evaluation was performed. The test was designed by following the heuristics proposed by Nielsen [33]. This test was conducted by several members of the authors' department who assessed the application by completing a checklist. This questionnaire focused on aspects such as identity and information, language, structure and navigation, layout, help elements, user feedback, and so forth. In general, the results provided us with valuable tips for improving the finally delivered version of the application. The results of this evaluation are also available on the developer's website [31].

4. Experts' Survey on Evidence-Based Systematic Review

The proposed framework collects all the main desirable features in a tool that supports the complete SR process. In this section, an expert survey for assessing the features included in the framework is presented. The results provide us with valuable insights for the improvement of the CloudSERA tool. The survey carried out involves the following steps: expert screening, survey questions' definition and implementation and, finally, data collection and analysis of the experts' opinions. In this study, the survey was conducted by 11 experts from different research academic disciplines, such as: Humanities, Applied Sciences, Formal Sciences, Social Sciences and Natural Sciences.

4.1. Expert Screening and Survey Questions

First, a purpose sampling technique was applied to select the experts from the sampling frame who had completed the survey. We considered researchers from different academic disciplines who had performed and supervised at least one SR. The expert screening, based on the previous characteristics, was carried out with researchers from the University of Cádiz INDESS Research Institute, a multidisciplinary research institute whose members' areas match the goal of the study. In addition, researchers from other universities also participated.

Second, the expert survey [34] was designed by following the recommendations provided by Oates [16] and was published using Google Forms. The survey content was organized into six sections. The first section included a question related to data protection issues. Users needed to provide consent to allow for the analysis of their data. We also provided users with a consent revocation form [35] if required. The second section included questions related to the user's profile, whereas the third was devoted to obtaining data on their level of expertise with systematic reviews. The goal of the fourth and fifth sections was to validate the features contemplated in the information management and evidence synthesis categories. Finally, the sixth section aimed to capture the users' interest level in the SR support tools and validate the most relevant features contemplated in the non-functional and overall functionality categories.

The survey included several types of questions such as scale questions, multiple selection questions, and open questions to answer with free text. In the multiple selection and scale questions, respondents were entitled to include alternative responses, which was useful for indicating, for example, additional features to consider in an SR tool besides those already considered by the framework. The questions included are listed in Appendix A.

Once the data form was designed, an e-mail with detailed instructions was sent to the screened experts to complete the survey.

4.2. Data Collection and Analysis

Table 4 summarizes the experts' opinions about the significance of each feature included in the analysis framework. In order to properly analyze the results, some dimensions of interest, namely academic discipline and expertise level, were considered.

For each dimension considered, the average score assigned to the questions by the experts pertaining to the indicated dimension has been included in the table. Additionally, the total average of the scores of the entire sample of experts has also been provided.

- The complete sample: a total of 11 experts were involved in the survey.
- Research academic discipline: Humanities (2), Applied sciences (2), Formal sciences (2), Social sciences (3) and Natural Sciences (2).
- Expertise level: researchers who have supervised more than one SR (8) and those who have only supervised one (3).

The data collected for each section of the survey is analyzed in detail below. In order to make the study reproducible, a spreadsheet with data collection and analysis is available online [36].

Table 4. Results of the survey by feature category.

Dimension	Information Management	Evidence Synthesis	Overall Functionality
Academic Discipline			
Humanities	3.58	3.36	3.17
Applied sciences	4.75	4.75	4.58
Formal sciences	3.5	2.82	2.91
Social sciences	4.28	4.31	3.5
Natural sciences	4.58	4.32	4.17
Expertise level			
Medium	3.22	4	3.94
High	3.81	4.21	3.88
Complete sample			
	3.65	4.15	3.9

4.2.1. Information Management Features

The responses collected for the questions belonging to the information management category are discussed below:

- First, this section of the survey asked the experts about the reference management systems they use. In this case, the experts' responses enumerated the following: Mendeley, Zotero, JabRef, EndNote, and RefWorks.
- Second, they were asked about the search engines used. In this case, the experts from the Humanities mentioned ACM Digital Library, JSTOR, Web of Science, Google Scholar, and ResearchGate. On the other hand, the experts from the Applied Sciences mentioned PubMed, Scopus, EBSCO CINAHL, and MathSciNet; and experts from the Social Sciences mentioned—in addition to the search engines used in Humanities—Medline, Wolters Kluwer Ovid, and Sociological Abstracts. Researchers from the Natural Sciences mentioned AGRIS, Agricola, Academic Search Premier, CAB Direct, GreenFILE, Aquatic Sciences and Fisheries Abstracts, PsycINFO, DOAJ, EconLit, Sociological Abstracts, ProQuest Dissertations and Theses, DART eTheses Portal, and EThOS.
- Third, the experts were asked about the guides used to measure the quality of the methodology used in their selected studies. In this case, only the experts from the Applied Sciences, Natural Sciences and Social Sciences use PRISMA [37], ENTREQ [38] and Cochrane Collaboration [39].

Then, the experts were asked about the significance score they assigned to the features included in the Information management category. Table 4 presents the results of the experts' opinions, classified by their research disciplines and expertise level. From these results, we can observe the following:

- All features included in this category were considered relevant for the experts.
- The duplicate deletion feature has a lower than average score; this might indicate that integrating the remaining features into the SR tool to cover the experts' requirements should be a requisite. Besides, this can indicate that the features needed for information management in the SR process are applied in the same way in all disciplines.
- The experts suggested including the following features: stakeholder engagement, inter-rater reliability in the screening process, detect future lines, new questions and challenges, and the possibility of setting the sample size.

4.2.2. Evidence Synthesis Features

The responses collected for the questions belonging to the evidence synthesis category are discussed below:

- First, this section asked the experts about the type of study more frequently conducted or supervised. In this case, four researchers indicated quantitative studies, two researchers indicated qualitative studies and five researchers indicated mixed studies.
- Second, they were asked about the techniques used to synthesize collected evidence. In this case, the researchers mentioned the following techniques: grounded theory, content analysis, case survey, meta-study, meta-ethnography, thematic analysis, narrative summary, Bayesian meta-analysis and meta-study. Additionally, they contributed qualitative comparative analysis method and meta-synthesis to the previous list.
- Third, the experts were asked about the method used to collect data from primary studies. In this case, the experts responded with: manually, using Mendeley, using online survey data, using face-to-face surveys or phone surveys. Other experts indicated that, in certain disciplines, it is complicated to gather data because “some studies do not provide data, for example, patient data are not commonly available”.
- Fourth, they were asked about tools used to analyze data. In this case, the experts from the Humanities mentioned Atlas.ti, or SPSS, experts from the Social Sciences mentioned R, meta-regression and Forest Plot, and experts from the Natural Sciences named Microsoft Excel. Other researchers indicated that they manually perform the analysis, without using any support tool.
- Fifth, the experts were asked about the techniques used to measure the risk of bias assessment. In this case, only the experts from the Applied Sciences and Social Sciences indicated that they used this type of technique. In the case of the Applied Sciences, experts performed a revision using a risk of bias table, whereas they used GRADE and AHQR guidelines in the case of the Social Sciences.
- Sixth, they were asked about the methods used for representing results. In this case, seven of them indicated that they used visual representations for synthesizing results; five of them indicated that they used a flowchart to depict the selection process of each study; and three of them indicated that they used visual representations for indicating the included and excluded studies.
- Seventh, the experts were asked about their opinion on whether quantitative reports should be different to qualitative ones. In this case, seven of them responded affirmatively and two of them negatively.
- Eighth, they were asked about including extra information in the reports. In this case, one expert from the Social Sciences indicated that the addition of summary tables and additional files with the complete information is needed and one expert from the Applied Sciences indicated that it is relevant to include personal opinions.

Then, the experts were asked about the relevance score they assigned to the features included in the evidence synthesis category. Table 4 shows the results of the experts' opinions. From the previous results, we can observe:

- All features included in this category were considered relevant by experts.
- The experts from the Humanities and Formal Sciences gave a higher score to the ES features than experts from the Applied Sciences, Social Sciences and Natural Sciences. This might indicate that the experts of the former disciplines invest a greater effort into the application of evidence synthesis techniques in their SR processes.

4.2.3. Overall Features

The responses collected for the questions belonging to the overall category are discussed below:

- First, the experts were asked about their interest level in the availability of tools to support the SR process. In this case, the average level of interest is 4.18, indicating that the inclusion of this kind of tool is very relevant.
- Second, they were asked whether they have used SR tools previously. In this case, only one of the participants had used this type of tool and mentioned EPPI-Reviewer, CADIMA, Rayyan, SysRev and Colandr.

Then, the experts were asked about the relevance score they assigned to the features included in the overall functionality category. Table 4 shows the results of the experts' opinions. From the previous results, we can also observe that all the features belonging to the overall functionality category were considered to be relevant by the experts. We can observe that the experts in each domain assigned quite different scores for each feature. However, they were all rated above 3, meaning that these features are relevant or very relevant for them in SR support tools.

5. Results

Following the opinions and analysis previously discussed, we can draw a road map for enriching CloudSERA with the most valuable features required to provide more complete support to the SR process.

- Provide a more exhaustive integration with other reference management systems, such as Zotero, JabRef, EndNote, and RefWorks.
- Extend the built-in search engines, including others such as: JSTOR, Web of Science, ResearchGate, PubMed, EBSCO CINAHL, MathSciNet, Medline, Ovid, Sociological Abstracts, AGRIS, Agricola, Academic Search Premier, CAB Direct, Aquatic Sciences and Fisheries Abstracts, GreenFILE, PsycINFO, DOAJ, EconLit, Sociological Abstracts, ProQuest Dissertations and Theses, DART eTheses Portal, and EThOS.
- Implement the integration with guides to measure the quality of the methodology used in the primary studies, such as: PRISMA, ENTREQ or Cochrane collaboration.
- Include computational support to partially automate some of the evidence data synthesis techniques, such as: grounded theory, content analysis, case survey, meta-ethnography, meta-study, narrative summary, meta-study, qualitative comparative analysis methods, meta-synthesis, and thematic analysis.
- Provide the following features from the information management category: Inter-rated reliability in the screening process, detect future lines, new questions and challenges, and the ability to set the sample size.

Finally, an analysis of threats to validity is required. In this case, we followed a detailed protocol to define the more desirable features in SR support tools and the expert survey, in accordance with internal validity and construct validity. In addition, we screened the experts ensuring that they have experience in the SR process by having supervised at least one SR. The sample size is not large, but it helped us to validate the features proposed in the evidence-based SR framework and to propose new ones, thus fulfilling the goal of the survey.

6. Conclusions and Future Work

Undertaking systematic reviews is an essential task before starting any research activity. Researchers need to perform a preliminary study of the literature in order to know the current state-of-the-art in a specific research topic. Manually performing this task is very time-consuming for a user. For this reason, several tools have been developed to support systematic reviews. CloudSERA is a cloud-based systematic review tool focused on the realization of systematic literature reviews within the domain of Computer Science. In this work, we have carried out an analysis aimed at gathering the essential evidence synthesis features to provide for a more complete evidence synthesis process in a general-purpose systematic review tool. We have defined a framework that incorporates these features to evaluate and compare existing tools in other domains. With the aim of evaluating the framework and feeding it with more relevant features, an expert survey was carried out. That study has provided us with valuable insights to help improve the CloudSERA tool in supporting the complete set of evidence synthesis functionalities. Moreover, after evaluating existing tools that perform SR processes, we conclude that none of them incorporate all the necessary functionalities for domain experts. For this reason, enhancing CloudSERA by integrating these functionalities can be a promising and viable option.

We plan to extend CloudSERA with an On-Line Analytical Processing (OLAP) viewer to carry out multi-dimensional analysis easily. In addition, data mining algorithms will be explored and included to discover and cluster data extracted from the primary studies. Additionally, we will incorporate a workflow engine to orchestrate the execution of the tasks to complete an SR process. Finally, we also plan to conduct a heuristic evaluation with the next CloudSERA version using potential end-users from different knowledge areas to measure usability attributes such as learnability, efficiency and user satisfaction.

Author Contributions: Conceptualization, I.R.-R.; methodology, T.P. and J.M.D.; validation, M.J.C. and A.T.; formal analysis, T.P. and J.M.M.; investigation, T.P.; data curation, T.P.; writing—original draft preparation, T.P. and I.R.-R.; writing—review and editing, J.M.D. and A.T.; visualization, J.M.M.; supervision, I.R.-R. and J.M.D.; project administration, J.M.D.; funding acquisition, J.M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Spanish Research Agency (Agencia Estatal de Investigación) with ERDF funds grant number TIN2017-85797-R (VISAIGLE project). The research stay of T. Person in the SFU was funded by Erasmus+ KA107 grant number 2017-1-ES01-KA107-037422. The APC was funded by the VISAIGLE project.

Data Availability Statement: The data used in this paper are available online: the experts' survey and consent form [34]; consent revocation form [35]; and data collection and analysis spreadsheet [36].

Acknowledgments: The authors acknowledge the work of Ángel Rafael González Toro in the software implementation of the first version of CloudSERA tool.

Conflicts of Interest: The authors declare that they have no conflict of interests.

Abbreviations

The following abbreviations are used in this manuscript:

ES	Evidence Synthesis
SR	Systematic Review
RQ	Research Question
SMS	Systematic Mapping Study
SLR	Systematic Literature Review
OLAP	On-Line Analytical Processing

Appendix A. Survey Questions

The questions included in the expert's survey are intended to collect the following data by category:

Appendix A.1. User Profile

- E-mail.
- Full name.
- Highest academic degree.
- Kind of research: academic research, industrial research, government research, etc.
- Academic discipline.

Appendix A.2. Systematic Review Experience

- Number of SRs performed.
- Number of SRs supervised.
- Level of expertise doing SRs (in a scale from 1 to 5).

Appendix A.3. Information Management

- Most frequently utilized reference management systems: EndNote, Mendeley, RefWorks, Zotero, CiteULike, JabRef, etc.
- Most frequently used search engines: Web of Science, Springer Link, Google Scholar, etc.

- Followed methodological guides: Cochrane Collaboration, Kitchenham's guidelines, etc.
- Personal rating of significance of each feature collected in the *Information management* category of the framework (in a scale from non-relevant to essential).

Appendix A.4. Evidence Synthesis

- Type of the study more frequently carried out or supervised: quantitative, qualitative, or both quantitative and qualitative.
- Used techniques for data synthesis: thematic analysis, meta-study or Bayesian meta-analysis, among others [8].
- Methods used to collect data from studies.
- Tools used to analyze studies' data.
- Techniques used to assess the risk of bias.
- Representation techniques used to convey the results of the review study: flowcharts with the selection of studies process, tables/charts with percentages of studies analyzed or discarded according to the inclusion/exclusion criteria, tables/charts summarizing the synthesized data of the primary studies, etc.
- Opinion about the differentiating factors between a qualitative report and a quantitative one.
- Additional information commonly included in the final reports.
- Personal rating of significance for each feature collected in the Evidence synthesis category of the framework (in a scale from non-relevant to essential).

Appendix A.5. Systematic Review Tools

- Interest level in the availability of tools to support the SR process.
- Tools used during the SR process.
- Personal rating of significance for each feature collected in the Non-functional and Overall functionality categories of the framework (scale from non-relevant to essential).

References

1. Moed, H.F.; Glänzel, W.; Schmoch, U. *Handbook of Quantitative Science and Technology Research*; Springer: Dordrecht, The Netherlands, 2015.
2. Börner, K.; Chen, C.; Boyack, K.W. Visualizing knowledge domains. *Annu. Rev. Inf. Sci. Technol.* **2003**, *37*, 179–255. [CrossRef]
3. Cobo, M.J.; López-Herrera, A.G.; Herrera-Viedma, E.; Herrera, F. Science mapping software tools: Review, analysis, and cooperative study among tools. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 1382–1402. [CrossRef]
4. Fortunato, S.; Bergstrom, C.T.; Börner, K.; Evans, J.A.; Helbing, D.; Milojević, S.; Petersen, A.M.; Radicchi, F.; Sinatra, R.; Uzzi, B.; et al. Science of science. *Science* **2018**, *359*, eaao0185. [CrossRef] [PubMed]
5. GRADEpro. GRADEpro Guideline Development Tool [Software]. McMaster University (Developed by Evidence Prime, Inc.). Available online: [grade.pro.org](https://www.gradepro.org) (accessed on 22 May 2021).
6. Noblit, G.W.; Hare, R.D.; Hare, R. *Meta-Ethnography: Synthesizing Qualitative Studies*; Sage: Newbury Park, London, UK, 1988; Volume 11.
7. Cooper, H.; Hedges, L.V.; Valentine, J.C. *The Handbook of Research Synthesis and Meta-Analysis*; Russell Sage Foundation: New York, NY, USA, 2009.
8. Dixon-Woods, M.; Agarwal, S.; Jones, D.; Young, B.; Sutton, A. Synthesising qualitative and quantitative evidence: A review of possible methods. *J. Health Serv. Res. Policy* **2005**, *10*, 45–53. [CrossRef] [PubMed]
9. Cruzes, D.S.; Dybå, T. Synthesizing evidence in software engineering research. In Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, Bolzano, Italy, 16–17 September 2010; p. 1.
10. Cochrane Collaboration. *The Cochrane Reviewers' Handbook Glossary*; Cochrane Collaboration: 2001. Available online: <https://books.google.es/books?id=xSUzHAAACAAJ> (accessed on 22 May 2021).
11. Kitchenham, B.A.; Budgen, D.; Brereton, P. *Evidence-Based Software Engineering and Systematic Reviews*; CRC Press: Boca Raton, FL, USA, 2015; Volume 4.
12. Daniels, K.; Langlois, É.V. Applying Synthesis Methods in Health Policy and Systems Research. In *Evidence Synthesis for Health Policy and Systems: A Methods Guide*; World Health Organization: Geneva, Switzerland, 2018; Chapter 3, pp. 26–39.
13. Haddaway, N.R.; Bilotta, G.S. Systematic reviews: Separating fact from fiction. *Environ. Int.* **2016**, *92–93*, 578–584. [CrossRef] [PubMed]

14. Kitchenham, B.; Charters, S. *Guidelines for performing Systematic Literature Reviews in Software Engineering*; Technical Report EBSE 2007-001, Keele University and Durham University Joint Report; 2007. Available online: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.471&rep=rep1&type=pdf> (accessed on 22 May 2021).
15. Petersen, K.; Vakkalanka, S.; Kuzniarz, L. Guidelines for conducting systematic mapping studies in software engineering: An update. *Inf. Softw. Technol.* **2015**, *64*, 1–18. [CrossRef]
16. Oates, B.J. *Researching Information Systems and Computing*; Sage: London, UK, 2005.
17. Ruiz-Rube, I.; Person, T.; Mota, J.M.; Doder, J.M.; González-Toro, Á.R. Evidence-Based Systematic Literature Reviews in the Cloud. In *Intelligent Data Engineering and Automated Learning—IDEAL 2018*; Yin, H., Camacho, D., Novais, P., Tallón-Ballesteros, A.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 105–112.
18. Kohl, C.; McIntosh, E.J.; Unger, S.; Haddaway, N.R.; Kecke, S.; Schiemann, J.; Wilhelm, R. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environ. Evid.* **2018**, *7*, 8. [CrossRef]
19. Hassler, E.; Carver, J.C.; Hale, D.; Al-Zubidy, A. Identification of SLR tool needs—results of a community workshop. *Inf. Softw. Technol.* **2016**, *70*, 122–129. [CrossRef]
20. Manterola, C.; Astudillo, P.; Arias, E.; Claros, N.; Mincir, G. Revisión sistemática de la literatura. Qué se debe saber acerca de ellas. *Cirugía Espa Nola* **2013**, *91*, 149–155. [CrossRef] [PubMed]
21. Cheng, S.; Augustin, C.; Bethel, A.; Gill, D.; Anzaroot, S.; Brun, J.; DeWilde, B.; Minnich, R.; Garside, R.; Masuda, Y.; et al. Using machine learning to advance synthesis and use of conservation and environmental evidence. *Conserv. Biol.* **2018**, *32*, 762–764. [CrossRef] [PubMed]
22. Tan, M.C. Colandr. *J. Can. Health Libr. Assoc. De L'Association Des Bibliothèques De La Santé Du Can.* **2018**, *39*, 85–88. [CrossRef]
23. Partners, E. *DistillerSR [Computer Program]*; Evidence Partners: Ottawa, ON, Canada, 2011.
24. Thomas, J.; Brunton, J.; Graziosi, S. EPPI-Reviewer 4.0: Software for Research Synthesis. EPPI-Centre Software. London: Social Science Research Unit. *Inst. Educ. Univ. Lond.* **2010**. Available online: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=1913> (accessed on 2 May 2021).
25. Lajeunesse, M.J. Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for R. *Methods Ecol. Evol.* **2016**, *7*, 323–330. [CrossRef]
26. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan—A web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 210. [CrossRef] [PubMed]
27. (ESEG), E.S.E.G. *REviewER*, 2013. Available online: <https://sites.google.com/site/eseportal/tools/reviewer> (accessed on 22 May 2021).
28. Molléri, J.S.; Benitti, F.B.V.; da Rocha Fernandes, A.M. Aplicação de Conceitos de Inteligência Artificial na Condução de Revisões Sistemáticas: Uma Resenha Crítica. *An. Sulcomp* **2013**. Available online: <http://periodicos.unesc.net/index.php/sulcomp/article/viewFile/1028/971> (accessed on 22 May 2021).
29. Fernández-Sáez, A.M.; Bocco, M.G.; Romero, F.P. SLR-Tool: A Tool for Performing Systematic Literature Reviews. *ICSOF* **2010**, *2*, 157–166.
30. González Toro, n.R.; SPI&FM, T. CloudSERA's Tool. 2019. Available online: <http://slr.uca.es/> (accessed on 22 May 2021).
31. González Toro, n.R.; SPI&FM, T. CloudSERA's GitHub Repository. 2019. Available online: <https://github.com/spi-fm/CloudSERA> (accessed on 22 May 2021).
32. Von Alan, R.H.; March, S.T.; Park, J.; Ram, S. Design science in information systems research. *MIS Q.* **2004**, *28*, 75–105.
33. Nielsen, J. Ten usability heuristics, 2005. Available online: <https://pdfs.semanticscholar.org/5f03/b251093aee730ab9772db2e1a8a7eb8522cb.pdf> (accessed on 22 May 2021).
34. Person, T.; Ruiz-Rube, I.; Mota, J.M.; Cobo, M.J.; Tselykh, A.; Doder, J.M. Expert Survey Form. 2019. Available online: <http://bit.ly/cloudera-experts-survey> (accessed on 22 May 2021).
35. Person, T.; Ruiz-Rube, I.; Mota, J.M.; Cobo, M.J.; Tselykh, A.; Doder, J.M. Consent Revocation Form for Data Collection and Analysis for Participants in the Expert Survey. 2019. Available online: <http://bit.ly/cloudera-revocate-form> (accessed on 22 May 2021).
36. Person, T.; Ruiz-Rube, I.; Mota, J.M.; Cobo, M.J.; Tselykh, A.; Doder, J.M. Spreadsheet with the Data Collected and Analyzed in the Survey. Available online: <http://bit.ly/cloudsera-data> (accessed on 22 May 2021).
37. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Group, P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef] [PubMed]
38. Tong, A.; Flemming, K.; McInnes, E.; Oliver, S.; Craig, J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med. Res. Methodol.* **2012**, *12*, 181. [CrossRef] [PubMed]
39. Godlee, F. The Cochrane collaboration. *Br. Med. J.* **1994**, *309*, 969. [CrossRef] [PubMed]