Digital Contact Tracing Applications for COVID-19: A Citizen-Centred Evaluation Framework

Damyanka Tsvyatkova, Manzar Abbas, Sarah Beecham, Jim Buckley, Muslim Chochlov, Brian Fitzgerald, Liam Glynn, Kevin Johnson, John Laffey, Bairbre McNicholas, Bashar Nuseibeh, Michael O'Callaghan, James O'Connell, Derek O'Keeffe, Ian R. O'Keeffe, Abdul Razzaq, Kaavya Rekanar, Ita Richardson, Andrew Simpkin, Jane Walsh, Thomas Welsh, Cristiano Storni

Submitted to: JMIR mHealth and uHealth
on: May 25, 2021

Table of Contents

Digital Contact Tracing Applications for COVID-19: A Citizen-Centred Evaluation Framework

Damyanka Tsvyatkova[1] PhD; Manzar Abbas[1] PhD; Sarah Beecham[1] PhD; Jim Buckley[1, 2] Prof Dr; Muslim Chochlov[1] PhD; Brian Fitzgerald[1] Prof Dr; Liam Glynn[3] Prof Dr; Kevin Johnson[4] Prof Dr; John Laffey[5] Prof Dr; Bairbre McNicholas[5] PhD; Bashar Nuseibeh[1] Prof Dr; Michael O'Callaghan[3] BMBS, BEng; James O'Connell[1]; Derek O'Keeffe[5] Prof Dr; Ian R. O'Keeffe[1] PhD; Abdul Razzaq[1] PhD; Kaavya Rekanar[1]; Ita Richardson[1, 2] Prof Dr; Andrew Simpkin[6] PhD; Jane Walsh[7] PhD; Thomas Welsh[1] PhD; Cristiano Storni[2] PhD

[1]Lero, Science Foundation Ireland Research Centre for Ireland University of Limerick Limerick IE
[2]Department of Computer Science and Information Systems University of Limerick Limerick IE
[3]School of Medicine University of Limerick Limerick IE
[4]Department of Nursing and Midwifery University of Limerick Limerick IE
[5]School of Medicine National University of Ireland Galway Galway IE
[6]School of Mathematics, Statistics and Applied Mathematics National University of Ireland Galway Galway IE
[7]School of Psychology National University of Ireland Galway Galway IE

Corresponding Author:
Cristiano Storni PhD
Department of Computer Science and Information Systems
University of Limerick
University of Limerick
CS2-032
Limerick
IE

Abstract

Background: The silent transmission of COVID-19 has led to an exponential growth of fatal infections. With over 3 million deaths world-wide, the need to control and stem transmission has never been more critical. New COVID-19 vaccines offer hope. However, administration timelines, long-term protection, and effectiveness against variants are still unknown. In this context, Contact Tracing, and digital Contact Tracing Apps (CTAs) continue to offer a mechanism to help contain transmission, keep people safe, and help kickstart economies. However, CTAs must address a wide range of often conflicting concerns which make their development/evolution complex: for example, the app must preserve citizens' privacy whilst gleaning their close contacts and as much epidemiological information as possible.

Objective: In this paper, we derive a compare-and-contrast evaluative framework for CTAs that captures best-of-breed development and evolution concerns for CTAs organized into seven pillars. As our goal is to integrate and expand on existing work in this domain, with a particular focus on citizen adoption, we call this framework the Citizen-Focused Compare-and-Contrast Evaluation Framework (C3EF) for CTAs.

Methods: The framework has been derived through mixed methods. First, we reviewed the related literature on CTAs and mHealth app evaluations, from which we derived a preliminary set of attributes and organizing pillars. These attributes were validated, augmented, and refined by applying the provisional framework against a selection of CTAs. At this point, questions to probe each attribute of the framework were formulated and iteratively tested on selected CTAs. Each framework pillar was then subjected to internal cross-team scrutiny where domain experts responsible for developing a pillar defended its sufficiency, relevancy, specificity, and non-redundancy. The consolidated framework was further validated on the selected CTAs to create a refined version of C3EF for CTAs.

Results: The final framework presents seven pillars exploring issues related to CTA's design, adoption, and use: (General) Characteristics, Usability, Data Protection, Effectiveness, Transparency, Technical Performance, and Citizen Autonomy. The pillars encompass attributes, sub-attributes, and a set of illustrative questions (with associated example answers) to support app design, evaluation, and evolution. An online version of the framework has been made available to developers, health authorities, and others interested in assessing CTAs.

Conclusions: Our CTA evaluation-concerns framework provides a holistic compare-and-contrast tool that supports the work of decision-makers in the development and evolution of CTAs for citizens. This framework supports reflection on design decisions to better understand and optimize the design compromises in play when evolving current CTAs for increased public adoption. We intend it to act as a foundation for other researchers to build on and extend, as the technology matures and new CTAs become available.

(JMIR Preprints 25/05/2021:30691)
DOI: https://doi.org/10.2196/preprints.30691

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Digital Contact Tracing Applications for COVID-19: A Citizen-Centred Evaluation Framework

Damyanka Tsvyatkova[1], Manzar Abbas[1], Sarah Beecham[1], Jim Buckley[1, 2], Muslim Chochlov[1], Brian Fitzgerald[1], Liam Glynn[3], Kevin Johnson[4], John Laffey[5], Bairbre McNicholas[5], Bashar Nuseibeh[1], Michael O'Callaghan[3], James O'Connell[1], Derek O'Keeffe[5], Ian R. O'Keeffe[1], Abdul Razzaq[1], Kaavya Rekanar[1], Ita Richardson[1, 2], Andrew Simpkin[6], Jane Walsh[7], Thomas Welsh[1], Cristiano Storni[2]

[1]Lero, Science Foundation Ireland Research Centre for Ireland, University of Limerick, Limerick, IE
[2]Department of Computer Science and Information Systems, University of Limerick, Limerick, IE
[3]School of Medicine, University of Limerick, Limerick, IE
[4]Department of Nursing and Midwifery, University of Limerick, Limerick, IE
[5]School of Medicine, National University of Ireland Galway, Galway, IE
[6]School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Galway, IE
[7]School of Psychology, National University of Ireland Galway, Galway, IE

Corresponding Author:
Cristiano Storni,
Department of Computer Science and Information Systems
University of Limerick
Limerick
IE

Abstract

Background: The silent transmission of COVID-19 has led to an exponential growth of fatal infections. With over 3 million deaths world-wide, the need to control and stem transmission has never been more critical. New COVID-19 vaccines offer hope. However, administration timelines, long-term protection, and effectiveness against variants are still unknown. In this context, Contact Tracing, and digital Contact Tracing Apps (CTAs) continue to offer a mechanism to help contain transmission, keep people safe, and help kickstart economies. However, CTAs must address a wide range of often conflicting concerns which make their development/evolution complex: for example, the app must preserve citizens' privacy whilst gleaning their close contacts and as much epidemiological information as possible.
Objectives: In this paper, we derive a compare-and-contrast evaluative framework for CTAs that captures best-of-breed development and evolution concerns for CTAs organized into seven pillars. As our goal is to integrate and expand on existing work in this domain, with a particular focus on citizen adoption, we call this framework the Citizen-Focused Compare-and-Contrast Evaluation Framework (C^3EF) for CTAs.
Methods: The framework has been derived through mixed methods. First, we reviewed the related literature on CTAs and mHealth app evaluations, from which we derived a preliminary set of attributes and organizing pillars. These attributes were validated, augmented, and refined by applying the provisional framework against a selection of CTAs. At this point, questions to probe each attribute of the framework were formulated and iteratively tested on selected CTAs. Each framework pillar was then subjected to internal cross-team scrutiny where domain experts responsible for developing a pillar defended its sufficiency, relevancy, specificity, and non-redundancy. The consolidated framework was further validated on the selected CTAs to create a refined version of C^3EF for CTAs.
Results: The final framework presents seven pillars exploring issues related to CTA's design,

adoption, and use: (General) Characteristics, Usability, Data Protection, Effectiveness, Transparency, Technical Performance, and Citizen Autonomy. The pillars encompass attributes, sub-attributes, and a set of illustrative questions (with associated example answers) to support app design, evaluation, and evolution. An online version of the framework has been made available to developers, health authorities, and others interested in assessing CTAs.

Conclusion: Our CTA evaluation-concerns framework provides a holistic compare-and-contrast tool that supports the work of decision-makers in the development and evolution of CTAs for citizens. This framework supports reflection on design decisions to better understand and optimize the design compromises in play when evolving current CTAs for increased public adoption. We intend it to act as a foundation for other researchers to build on and extend, as the technology matures and new CTAs become available.

Keywords:
COVID-19; mHealth; Digital Contact Tracing Apps; Framework; Evaluation

Introduction

The global coronavirus pandemic (COVID-19) calls for rapid measures to monitor and control the spread of the virus. Contact tracing is one of the measures adopted by health authorities. This approach has already been used with a certain level of success for other dangerous illnesses such as Tuberculosis [1] and Ebola [2]. As part of the contact tracing effort in the COVID-19 pandemic, the deployment of mobile apps and their potential in collecting, storing, and sharing citizens' contact tracing data has been examined, with early studies showing favorable results [3,4]. These studies have contributed to the impetus for using digital Contact Tracing Apps (CTAs), and many CTAs have been developed for nation's use, to facilitate community-based disease-surveillance [5].

The application of CTAs in real-world settings has provoked numerous discussions regarding their design [6–9], concerns about security, privacy of contact tracing app data, and the barriers for their widespread acceptance and adoption by citizens [10,11]. Reflecting these discussions, CTAs evaluation frameworks have emerged that specifically focus on different aspects, such as the assessment of contact tracing architectures [12], socio-technical issues [13], privacy [14,15], ethical and legal challenges [16], feasibility and effectiveness [17], usability [18] and essential attributes [19,20].

In this context, a legitimate concern is for a more comprehensive evaluation framework that would encompass a variety of aspects of CTAs, pertinent to the adopting citizens, and which would enable decision-makers (e.g. developers, health authorities) to assess and possibly improve their designs. This concern drives our research question, which is concerned with the need to devise an organized framework to enable a more comprehensive assessment of current CTAs, potentially increasing adoption. In this paper, we address this question by proposing a Citizen-Focused Compare-and-Contrast Evaluation Framework for CTAs (C^{3}EF). This holistically brings together and stress-tests existing, fragmented works on evaluation concerns for CTAs/mHealth applications. It aims to help in the assessment and improvement of existing CTAs solutions through a taxonomy of 7 pillars: (General) Characteristics, Usability, Data Protection, Effectiveness, Transparency, Performance, and Citizen Autonomy. This article introduces the C^{3}EF and presents its derivation over several iterations, to the proposed attributes and associated questions. It includes illustrations of the framework's application to a selection of existing CTAs.

The next section presents an overview of the method we used to develop C[3]EF. This is followed by the *C[3]EF framework* itself, which offers an overview and illustration of the resultant framework. Our *Conclusion* presents a summary of our contribution and future work.

Methods

Review and Framework Derivation

To accommodate the high-complexity and multi-disciplinary nature of CTAs evaluation for wide societal adoption, we used a multi-stage, mixed-methods approach, based on iterative refinement. Within this approach, the multi-disciplinary nature of the framework was specifically handled by 'segmentation' where domain experts within the research team were allocated responsibility for individual parts of the framework.

This methodology is portrayed in Figure 1 and was broken into three main phases over the period of 8 months – Phase 1: Derivation, Phase 2: Concretization and Critique, and Phase 3: Refinement and Finalization.

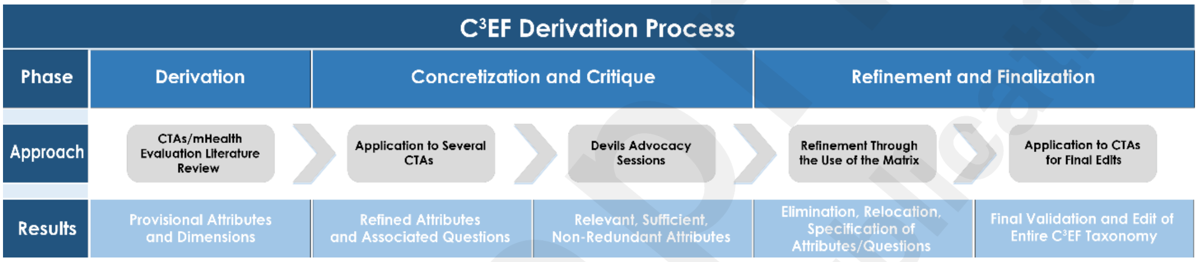


Figure 1. Activity phases and deliverables in the development of our Citizen-Focused Compare-and-Contrast Evaluation Framework for CTAs (C[3]EF) for CTAs.

Phase 1: Derivation

This phase was based on a literature review of relevant areas. To obtain a holistic perspective, we focused on existing frameworks (their attributes and characteristics) and on various guidelines and regulations for CTAs and mHealth apps in general. We included sources on:

1. CTAs: These included peer-reviewed literature on existing evaluation frameworks, grey publications discussing design characteristics and functionality, and design guidelines/EU regulations for development of effective digital CTAs.
2. Mobile applications: This included the most recent evaluation frameworks for mobile apps, evaluation frameworks for mHealth applications, accessibility principles for mobile apps, Universal Design (UD) for the apps, and taxonomies of usability.

The search itself was composed of three different strategies:

1. Electronic databases were leveraged to search academic texts: Google Scholar, Elsevier, ACM digital library, Sage, IEEE Xplore, and Springer.
2. Searches of web-based grey literature (e.g. via Google).
3. Consulting the reference lists of the selected articles to identify further relevant studies and using those sources to recursively increase our existing set of articles [21].

The search string was 'evaluation frameworks' AND 'digital contact tracing applications' AND 'COVID-19' OR 'mobile applications' OR 'mHealth applications' OR 'accessibility' OR 'universal design' OR 'usability' OR 'taxonomies' OR 'Data protection' OR 'GDPR' OR 'security threats'. Articles written in English and published between 2010 and 2020 were reviewed. Relevance was decided by reading the abstract, and, in cases where the abstract was insufficient, by reading their introduction and conclusion.

With this search strategy, 36 relevant sources (Multimedia Appendix 1) were identified. Twenty-one of these were distinct frameworks, 9 provided regulations and guidelines, and 6 others described important characteristics for CTAs. From these sources, an initial set of 104 attributes were manually extracted and were employed in the design of the original framework (Table 1).

Table 1. Identified attributes.

Sources	Attributes
CTAs specifically [6,7,9,13–17,22–30]	Purpose, Usability, Information Accuracy, Organizational Attributes/Reputation, Transparency, Privacy, User Control/Self-Determination, Approved by National Authority, Legal Compliance, Public Ownership, Reporting (Positive COVID-19 Test), Voluntary, Scientific Validity, Non-discrimination, Repurposing, Systematic Accountability, Public Benefit, Expiration (Limited Purpose), Scalability, Data Collection, Backend Server, Preserved Users' Anonymity, Data Protection, Data Management, Data Minimisation, Data Time Limitation, Securely Data Processing, Easy to Deactivate/Remove, Open-Source Code, Reliability, Explainability, Inclusiveness, Digital Inequality, Centralized, Decentralized, Alert Notification of Contacts, Using Google/Apple API, Not Using Google/Apple API, Bluetooth Technology, GPS- phone Location Data, DP-3T, WiFi-based Location Tracking
Mobile and mHealth Applications, UD for mHealth Apps for Older Individuals, Accessibility, Taxonomy, Interactive Technologies for Children, GDPR Regulations [31–49]	Effectiveness, Learnability, Memorability, Cognitive Load, Simplicity, Universality, Aesthetics, Security, Usefulness, Resources, Troubleshooting, Ongoing App Evaluation, Name of the Application, Navigation, Affordances, Interaction, Equitable Use, Flexibility, Ease of Use, Errors, Low Physical Effort, Size and Space for Approach and Use, Perceivable, Operability, Understandability, Robustness, Design, System, Content, Attractiveness, Comprehensibility, Accessibility, Consistency, Training, Trust, Battery Consumption, Less Storage Consumption, Adaptability, Performance, Layout, Platform Dependency, Onboarding, Speed, ISO 9241-11 (Efficiency, Satisfaction and Effectiveness), Parental/Legal Guardian Consent, Knowability, Clarity of the Elements, Clarity of the Structure, Clarity in Functioning, Helpfulness, Suitability of Documentation Content, Interactivity of Assistance, Completeness, Precision, Cultural Universality, Configurability, Workflow, Efficiency in Human Effort, Efficiency in Task Execution Time, Efficiency in Tied Up Efficiency in Economic Costs, Safety, Subjective Satisfaction, Interest

From an analysis of these attributes and from explicit classifications of attributes in several of the papers reviewed [13,14 ,17,22–24,29–31,38,41–43], we identified 6 evaluation pillars: *Usability*, *Data Protection*, *Effectiveness*, *Transparency*, *Technical Performance,* and the degree of *Autonomy* the app provides to downloading citizens. To uniquely identify an app and report on its non-evaluative characteristics, '*General Characteristics*' was also added. At this stage, the project team was divided into domain expert subgroups, one for each pillar, working on their specific development and further refinement. Overall, these subgroups reflected a range of competencies such as software engineering, Human-Computer Interaction (HCI), security, and data protection.

Usability

Usability refers to the ability of the CTAs to be easy to use and understand. We prioritized concerns of usability as the project-centered around increasing adoption by citizens, and therefore understanding its usability for target audiences was essential. We derived the initial attributes after a review of CTA evaluation frameworks [14–17,29] and from usability frameworks for mobile applications, mobile health applications (e.g. mHealth), Accessibility, and Universal Design (UD) literature [31,32,41,43–49]. Other sources that informed our deliberations were those discussing usability standards [50–52] in general, and EU design requirements [37]. Accessibility was included as a high-level attribute under usability, and it was mostly derived from the EU directive: Accessibility EN 301 549 [38], from where we took initial requirements from the directive and checked them against the Web Content Accessibility Guidelines [39] to formulate the required questions. An early report of the work in this pillar is available here [53].

Data Protection

Data Protection was chosen to accommodate societal concerns of privacy and security inspired by similar attributes in related works [42]. Although it has been noted for its complexity [54], we selected GDPR as our reference for the development of the Data Protection pillar: The General Data Protection Regulations (GDPR) of the European Union (EU) [55] is currently considered the foremost data protection legislation world-wide for protecting the rights of the individual. We retrieved the initial attributes and preliminary questions from national legal interpretations [56,57] for data-focused concerns. But such an approach excludes wider organizational attributes now found, for example, within our Transparency pillar. It also excludes system-oriented goals such as information security. Consequently, within this Data Protection pillar, we also developed a novel risk-based approach to comparing the system security of CTAs, based a review of related literature in mobile app GDPR evaluation [54,58–65].

Effectiveness

Effectiveness measures how successful an app is in terms of the accuracy of its contact tracing results, the COVID-restraining impact of the app over a jurisdiction, and the apps' popularity with citizens. Concerns include detecting and sharing close contacts, providing relevant information to citizens, and assessing their reactions to that information. We mainly employed the empirical guidelines presented by Wohlin and colleagues [21] regarding the evaluation of software tools, to this end. To specifically decide attributes relevant to the assessment of CTAs, we mapped these guidelines [21] to Vokinger's framework [29].

Transparency

While transparency is officially a subset of GDPR, a separate Transparency pillar was created in the framework to consider wider aspects of transparency not specifically related to functionality. For example, while the GDPR approach to transparency considers specific data-stores (such as locally stored contact information), transparency concerns such as the availability of a privacy policy or the open-sourcing of the source-code would not fit into that approach. Transparency here then concerns how open the developing organization is about its internal processes and artifacts. The initial attributes and their questions were formulated by extending our interpretation of GDPR [55], and considering already existing taxonomies in articles like [23,42].

Technical Performance

The Technical Performance pillar captures the efficiency of COVID-19 tracing applications: how efficient an application is in performing its functions. Particularly, the Technical Performance pillar focuses on system resource utilization and execution speed. The relevance of this pillar can be seen in how, for example, battery issues with the Exposure Notification Service provided by Google [66] and incorporated into the national Contact Tracker app in Ireland, caused battery issues over only one weekend and caused a large fall-off in app retention by the public [67]. The attributes for the Technical Performance pillar can be divided into resource utilization-related (e.g. CPU/disk/memory usage) and efficiency-related (e.g. response time). Because COVID-19 tracing applications are usually complex software systems (with dedicated front-end and back-end subsystems), the attributes can be applied to both subsystems respectively. The initial attributes for this pillar were derived from the “Performance Efficiency” category of ISO/IEC 25010: a software engineering quality model. The model is a standard for assessing characteristics of software systems and is widely applicable in software engineering.

Citizen Autonomy

Citizen Autonomy focuses upon the citizen’s ability to consent and the voluntary nature of the app. Its inclusion was inspired by the work of Gasser et al., 2020’s [16] study of digital tool’s ethical challenges. It was also based on the ‘User control/self-determination’ domain in Vokinger’s assessment framework for (COVID-19) CTAs [29] and the ‘autonomy’ category in [42]. In these works, the authors focused on users’ (existing) ‘data protection’ concerns which are mostly handled by our Data Protection pillar, but we wanted to extend the scope to specifically cover *initial* data access. Hence, the pillar focuses on a series of specific attributes that assess citizens’ control over the app’s access to phone functionalities like the camera, microphone, and GPS.

(General) Characteristics

‘General Characteristics’ refers to those characteristics which are non-evaluative, but which serve to distinguish the app from others and other versions of the app. Thus, the static information captured by the *Characteristics* pillar acts as a necessary first step to conducting the more in-depth compare and contrast evaluation found in the other pillars. An initial set of distinguishing characteristics were derived by examining three CTAs: SwissCovid (Switzerland) [68], Apturi Covid (Latvia) [69], and Immuni (Italy) [70], and related data retrieved from their AppStore, Google Play, and app home websites. Next, we analyzed the Google and Apple Exposure Notification (GAEN) API/framework [66] made available for use on Apple and Android devices, and the DP-3T protocol [71,72] that inspired Google’s API. We expanded our list of attributes further through a review of contact tracing protocols and frameworks listed on the Wikipedia COVID-19 Apps page [73]. Finally, we incorporated the literature review of app/mHealth-app evaluations (see Multimedia Appendix 2).

At the end of the work described above, our initial list of 104 attributes had grown to 139.

Phase 2: Concretization and Critique

In the second phase of development, our provisional framework was cross-checked against five CTAs which could be downloaded and activated in Ireland:

- HSE COVID Tracker app (Ireland) [74]

- PathCheck SafePlaces (MIT) (US) [75]
- NOVID (US) [76]
- Corona-Warn (Germany) [77], and
- Aman (Jordan) [78]

The two core considerations were to assess if the attribute could produce useful evaluation information on the CTA and how that information could be feasibly obtained. In this phase, feasibility concerns were sometimes overridden by the perceived importance of the information provided when probing a specific attribute. For example, the 'number of people alerted-early to their close contact status, who then go for testing', seems like one core 'Effectiveness' measure for CTAs. However, identifying this number involves a much wider information gathering and reporting effort than is normally available from the app itself, and so could be quite difficult to assess (e.g. members of the public would need to inform authorities when they turn up for testing that they have done so based on an alert issued to them by the app and that would need to be recorded on another system).

These apps reflected a *broad range* of approaches, as illustrated by their different *lead-bodies*, and the different *data protection philosophies* underpinning them. In terms of 'broad-range', the Republic of Ireland's app has provided the basis for apps in other jurisdictions, both in Europe and the United States [67]. In terms of 'lead bodies', these apps come predominantly from national health services, but PathCheck SafePlaces is an MIT-led initiative [79], and Novid is crowd-sourced, originating from CMU [76]. In terms of *data protection philosophies*, two of these apps originate in GDPR jurisdictions, but two originate from the United States, and one originates from Jordan.

The process followed in this phase was that the domain experts would apply their pillar to the five chosen apps to stress-test the ability of the framework to identify criticalities and key differences among apps. In addition, they formulated appropriate questions for each of the identified attributes and assessed the answers obtained to see if the attributes and the related questions had evaluative merit. Attributes were added where necessary. For example, in *General Characteristics*, the version number was identified as an important identifier, as CTAs, like other apps, tend to receive regular updates. Likewise, in *Effectiveness*, the effort/speed with which close contacts are alerted by CTAs was also identified as important. This concretization was sometimes complemented and reinforced with further, targeted reviews of literature where deemed necessary.

Devils' Advocate Sessions

A series of focused meets were held, where domain experts defended their redeveloped pillar in a Devil's Advocate session with the other team members. During these meetings, the team members dissected the pillar and tried to find fault with its attributes, under the headings of:

- *Relevance and Sufficiency*, where the team were encouraged to ask questions like 'why is this important?' and 'what else might be important?';
- *Specificity*, where domain experts were encouraged to hypothetically answer each of the associated questions in the pillar and to (thus) probe it for any ambiguity;
- *Cross-checking*, with their own pillars with others to identify possible overlaps in the framework.

The Devil's advocate sessions highlighted a number of changes mostly concerned with clarifying overlaps, identifying redundancies, and improving the structure and questions.

At the end of the grounding and critiquing exercises, we ended up with a total of 163 'grounded' attributes, and an initial formulation of 199 related questions. Additionally, we identified some cases to be discussed by the entire team during the third and final phase of our development.

Phase 3: Refinement and Finalization

Refinement

To support the refinement work of the entire team (especially in consideration of the remote work environment demanded by COVID), we created a Cross-pillar/ambiguity analysis matrix (Multimedia Appendix 3) which listed the ordered attributes and sub-attributes for each of the 7 pillars, assigning a unique identifier to each attribute, and a color code to each pillar. To support understanding of the attributes and to clarify their evaluative merit, we included questions and sample answers (based on our grounding exercise). We also noted (relevancy, sufficiency, specificity, and redundancy) issues for each attribute.

This structured format further supported domain experts in identifying outstanding similarities, overlaps, and issues of understandability. For instance, the framework probes *Citizen Autonomy* in terms of whether there is an official discussion forum for citizens using the app and whether that forum can be used to prompt change (CA01, CA02) (see Multimedia Appendix 3). It was noted that these overlap with attributes C16, a *Characteristic* attribute which probed the form of technical support, U73, a *Usability* attribute probing the existence of interactive assistance for technical support or any other mechanism to submit feedback on technical issues, bugs, and errors detected. A reorganization was proposed, deleting CA11 (redundant with CA01 and CA02), changing C16 from '*Does the app offer technical support?*', into '*What form of technical support is available for the users, to include synchronous and asynchronous support?*', and simplifying CA02 by removing its reference to *any other mechanism (to obtain technical support)*, as this was covered by the new phrasing of C16.

We had a total of 15 refinement sessions, with identified changes were included in the Cross-pillar/ambiguity analysis matrix. At the end of refinement, our framework was restructured to 161 attributes and 180 related questions (with sample answers), which now had internal consistency. Graphical visualizations of the refined pillars and their structures were also generated (Multimedia Appendix 4).

Finalization

Finally, the framework was applied to the newest versions of two of the apps employed in the 'grounding' phase to systematically double-check all attributes and questions so we could either confirm or implement a final edit. The main goal was to make sure questions are clear and understandable.

- HSE COVID Tracker app (Ireland) [74]
- NOVID (US) [76]

A number of other apps were also assessed less systematically, suggesting a number of minor edits to create a final version of the C^{3}EF for CTAs. The consolidated framework (with 7 pillars and 161 attributes) was presented for feedback to medical researchers and practitioners from the wider Covigilant group, the (Irish) Department of Health, and the (Irish) Health Service Executive's 'App Advisory Group'.

At this stage, we also created a web-based application [80] to make our framework available in the form of an online survey. This was devised to act as a demonstrator of our framework but also to possibly assist relevant stakeholders of CTAs in independently evaluating their work. This online tool offers visual overviews of the framework and gives access to the entire framework. With the depth and range of questions included in our C^{3}EF, the evaluation process may appear daunting and time-consuming. Consequently, we decided to provide access to individual pillars to enabling breaking down the assessment and allow stakeholders to select and prioritize their own assessment focus.

The C^{3}EF Framework

In this description of our final framework, we define each pillar and provide an overview of its specific attributes, sub-attributes, and questions. We then offer a selection of sample questions and answers to illustrate how we used the framework to evaluate, compare and contrast CTAs. (General) Characteristics are presented first in this instance, as they provide important contextualization/identification information for the other six pillars.

Characteristics

Characteristics captures non-judgmental criteria and factual information to identify and differentiate a given app and its main functionalities. App Characteristics are organized according to four headings: *General Characteristics, Availability, Organizational Reputation,* and *App Content* (Figure 2). Under these headings are a total of 25 specific questions that elaborate these app characteristics, all of which can be answered by direct inspection of the working app, through the information available on Google Play and Apple Store, or through the Developer website.

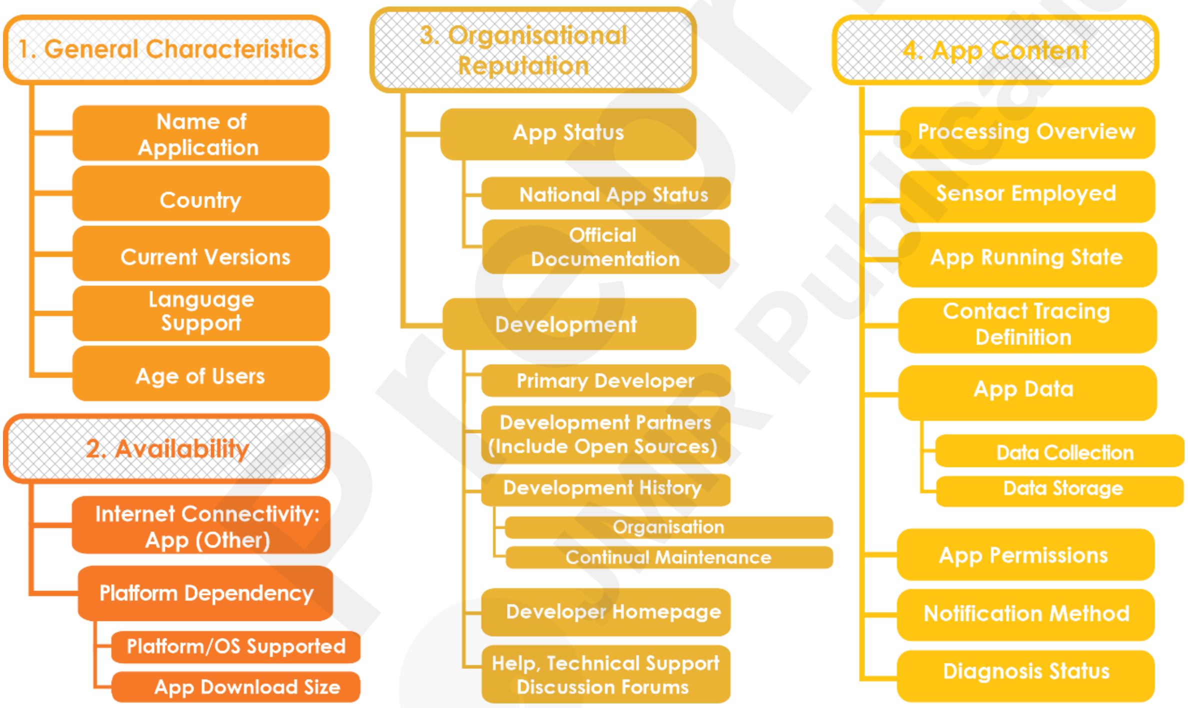


Figure 2. The Characteristics pillar.

General Characteristics captures four high-level app attributes, including the name of the app, the country it is developed for, the current versions of the app, the language(s) supported, and the age profile of the users, as advised on the downloading page.

Availability looks at Connectivity and Platform Dependency. We question whether an internet connection is needed to use the app. This is relevant as some apps require an internet connection even if they do not use the Internet or location-based information for their contact tracing (e.g. Jordanian app AMAN [78]). Questions considering what platforms (Android/iOS) are supported and the download size of the app also talk to the requirements for successfully executing the app.

Organizational Reputation looks at the status of the app, including whether the app is national and what official documentation is available. It also considers the organization which developed the app and whether any third-party or partners are involved in its development. It examines history of development, through an examination of the developers' prior experience of developing data-sensitive apps along with evidence of updates, enhancements, and maintenance of the actual product. Finally, the ability for users to ask questions and seek technical support is also probed.

While some aspects of organizational reputation (such as history) may disadvantage new startups, it is likely to play a key role in the confidence users have in app adoption. There is however, a level playing field when it comes to being transparent about the developers behind the development, and how the app is maintained.

App Content refers to what the app includes in terms of functionality and management of information. Definitions of contact tracing are queried, and information is given as to when and how contact tracing notifications are managed, since this may vary from country to country, or even app to app. For instance, the HSE app [74] states that the close contact notifications will be activated when there is 'direct exposure' to a positive case, where 'direct exposure' refers to 'within two metres for 15 minutes or more'. In contrast, NOVID [76] notifies users when other NOVID users close in their social network are positive cases. NOVID classifies proximity as 'near' or 'far', with six feet (2 metres) or under being 'near', and 12 feet (4 metres) or over being 'far'. App content also queries the permissions required by the app, to assess the information accessed.

Usability

Five high-level attributes have been identified under *Usability: Subjective Satisfaction, Universality, Design Effectiveness, User Interaction,* and *Ongoing App Evaluation* (Figure 3). Together they offer the opportunity to ask 86 specific questions about usability aspects of CTAs.

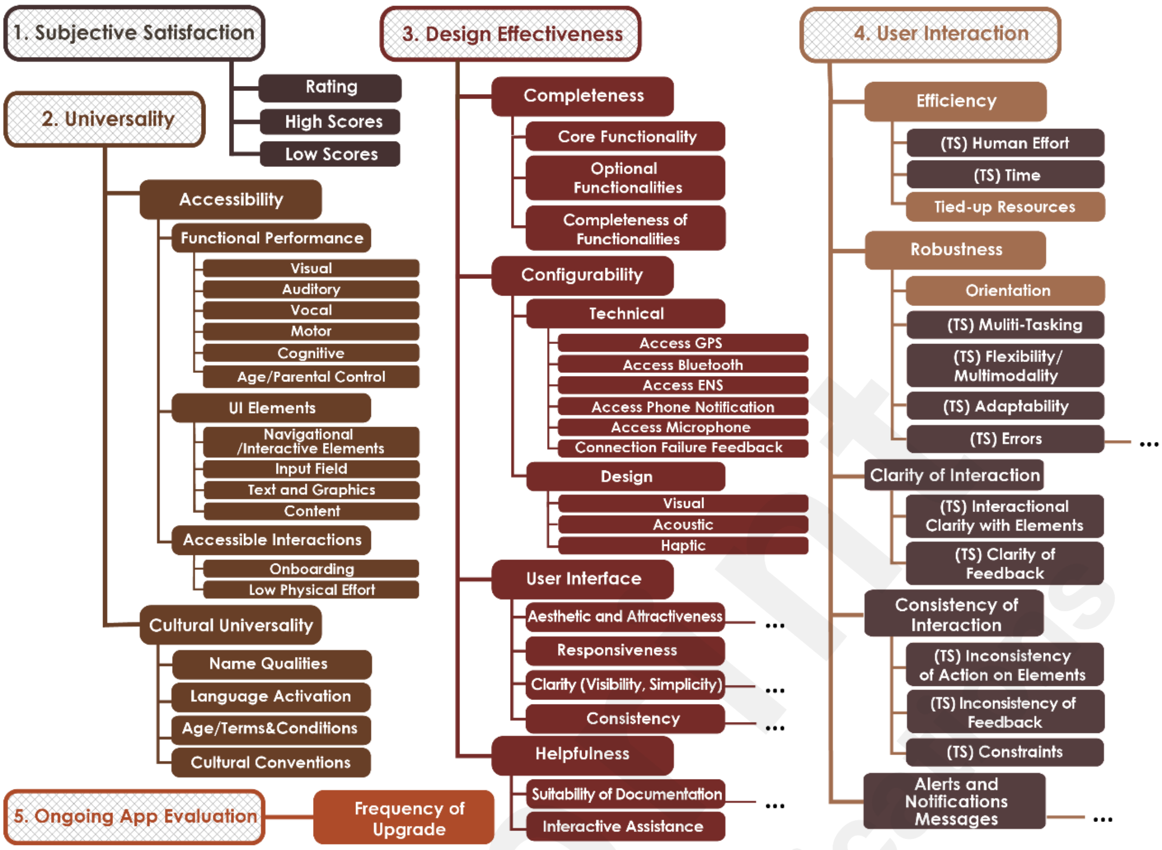


Figure 3. The Usability pillar.

Subjective Satisfaction looks at the perceived level of comfort experienced in using the application. As user retention is important in CTAs, this attribute looks at how citizens rate their experience in using the app. It includes a rating attribute (1 to 5), and two attributes looking at motivations for high or low scores. We typically look at the Google Play and Apple Store to inform our answers to these questions, though satisfaction could be better captured with longitudinal surveys.

Universality addresses population penetration, which is also key in the successful implementation of a CTAs strategy. Specifically, it aims to capture the ability of the app to be used by a variety of different users: users with potential impairments (physical or mental), but also users with different cultures/levels of education, or of different ages. The first is captured by the sub-attribute Accessibility, the latter by Cultural Universality. Accessibility refers to the quality of being 'easy to reach and use', and it mostly refers to users who might have a form of disability, impairment or limitation (either mental or physical). We covered three aspects related to Accessibility: Functional Performance, User Interface Elements, and Accessible Interactions. User Interface Elements look at the interface and how its elements adhere to general accessibility guidelines and EU regulations [6,38,39,43–45]. Accessible Interaction, which is the last sub-attribute under Accessibility, covers aspects such as Onboarding – features helping new users understand what the app does and learning how to use it, and the design of interactive elements to support Low Physical Effort (e.g. complete a task without scrolling, one-hand use, radio buttons). Cultural Universality is also part of Universality, and it focuses on the extent to which the system can be used by different users regardless of their cultural background and beliefs. We developed attributes and questions to cover aspects such as (a) availability of different language(s), (b) meanings that are evoked by the name of the app, (c) information on the age groups that the app targets, usually described in the 'Terms and Conditions' (see Figure 4, Example 1), and (d) design elements such as logos, colors, national flags, symbols for expressing cultural conventions.

Design Effectiveness covers several aspects concerning the capacity of the system, user interface, and interaction design of providing citizens with the necessary functionalities, options, commands, and supports. It includes four dimensions that are seen as key to convey correct utilization of the system and its adaptation to different contexts of use and user preferences: Completeness, Configurability, User Interface, and Helpfulness. Completeness was formulated to identify both essential and optional functionalities offered by the app and those features that are not included in the app, but that users (e.g. as voiced on online reviews) would like to have. Identification of core and optional functionalities help to also identify user tasks as the basis for task-specific questions, which will come later. Configurability looks at the capacity of the system to be personalized and to adapt to different contexts of use or users. It can be either technical (e.g. allowing independent activation or deactivation of GPS, Bluetooth, and other technologies) or related to the design (e.g. allowing personalization of visual, acoustic, and haptic feedback). User Interface deals with the assessment of the design elements used in the user interface. It includes four attributes: Aesthetic and Attractiveness (concerned with the Look'n'feel, color palette, and the name of the app); Responsiveness (concerned with the ability to adapt to different phone models and OSs); Clarity and Consistency of the design elements. These two last attributes offer Element Specific (ES) sub-attributes, in that they refer to and evaluate specific elements of the interface and not the app as a whole. An element can vary from a button to a menu, a slider or a table. A full list and a glossary of User Interface elements can be found here [81]). Our questions explore their perceptual and conceptual clarity, looking at their visibility, understandability (Figure 4, Example 2), and consistency (again both in terms of how their look and that their meaning is consistent throughout the whole app). But both Clarity and Consistency also apply to structures of elements (e.g. a button-bar containing buttons) in terms of how elements are logically grouped and whether these logic-based-groupings are consistent. The last sub-attribute of Design Effectiveness is Helpfulness looking at the Suitability of Documentation available to use and understand the app, and whether Interactive Assistance is available. Sub-attributes look at the availability of supportive information like Definitions of the terms used (e.g. what counts as a contact), Descriptions (see Figure 4, Example 3), Examples (e.g. tutorial, walk-through or explanatory videos showing how to use the app), and the availability of Interactive Assistance for troubleshooting (e.g. chatbots).

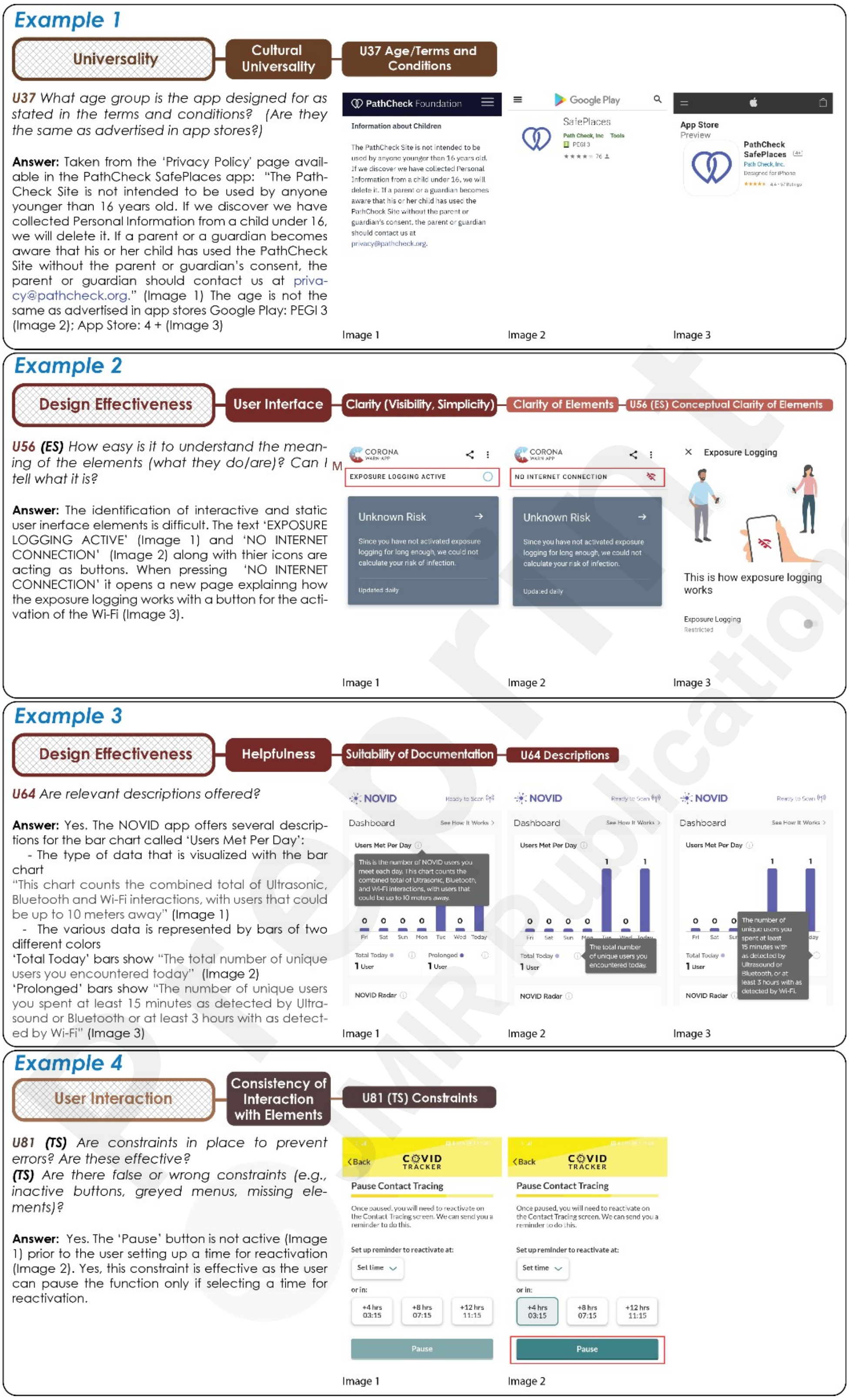


Figure 4. Example 1- PathCheck SafePlaces (US) [75]: age in 'Terms of Use'. Example 2 - Corona-Warn (Germany) [77]: understandability of interface elements. Example 3 - NOVID app (US) [76]: description. Example 4 - COVID Tracker app (Ireland) [74]: constraints.

User Interaction helps to assess the user's interaction with the application interface in the execution of specific tasks. Like Design Effectiveness, user interaction is important to ensure correct use of the CTA and lessening user frustration, but while design effectiveness looks at what the interface statically offers, looks, and conveys (e.g. afford), user interaction looks at how the interface behaves when users interact with it (e.g. feedback). It includes five sub-attributes: Efficiency, Robustness, Clarity of Interaction with Elements, Consistency of Interaction with Elements, and Alerts and Notifications Messages. Most of the selected sub-attributes and their related questions are Task-Specific (TS): similar to Element-specific attributes, Task-Specific attributes and questions are specific to users carrying out one specific task from beginning to end (e.g. activating/de-activating the contact tracing functionality). Therefore, to answer TS questions, app inspection is needed. Efficiency explores the capacity of the system to produce appropriate results in return for the resources that are invested. Here we considered three elements: Human Effort – the number of steps that are needed to carry out a core task, Time – needed to perform that task and the Tied-up Resources – the potential need for external resources (e.g. power or Internet) to perform the task. Robustness deals with the capacity of the system to adapt to different user preferences and contexts of use, but also its ability to deal with user errors. With our attributes, we look at: landscape/portrait mode, Multi-Tasking when using Bluetooth, GPS, technologies for more than one app/task (e.g. while using Bluetooth earphones), and the availability of multiple ways to achieve task execution (e.g. shortcuts). Adaptability looks at supporting task execution in different environments (e.g. in the dark), while Errors looks at errors messages available in the app as a result of inappropriate interaction and the availability of error recovery options (e.g. undo, redo)/the reversibility of user actions. Clarity of Interaction with Elements is concerned with the clarity of what can be done with the elements available in the app interface (namely their affordances) and what happens when users interact with these elements (e.g. with respect to clarity/confusion of app feedback). Consistency of Interaction with Elements is next and looks at potential consistencies of Actions across the elements, the Inconsistency of Feedback, and the use of design Constraints (if any) to prevent human errors/guide toward correct use (Figure 4, Example 4). Alerts and Notifications Messages is the last sub-attribute of the user interaction attribute, and it refers to the alert messages and notifications used in the application. We included attributes and questions to assess various types of Alert Messages used in the app and to assess the availability of Notifications Control, particularly for built-in notification settings in the app and Notification Messages that alert the user who has been in close contact with someone that reported positive COVID-19 test. This part closes with a question concerning the ability of the user to access and perhaps even manipulate or visualize the generated contact tracing data (e.g. number of contacts in a day, week, etc.).

Ongoing App Evaluation refers to the app's maintenance and upgrading. This is included as it becomes evident in our study how app maintenance and periodical updates from developers are important for the CTA's ability to maintain retention and correct use by addressing issues (technical or related to the design), while also targeting new emerging needs, etc. It includes only one sub-attribute looking at the Frequency of Upgrade: it can be found in the Google Play and App Stores, where the app can be downloaded and installed.

Data Protection

The Data Protection pillar consists of two subcategories: *Security and GDPR*. The GDPR category focuses upon the rights of the individual citizens, while the security category takes a more data-centric view (Figure 5).

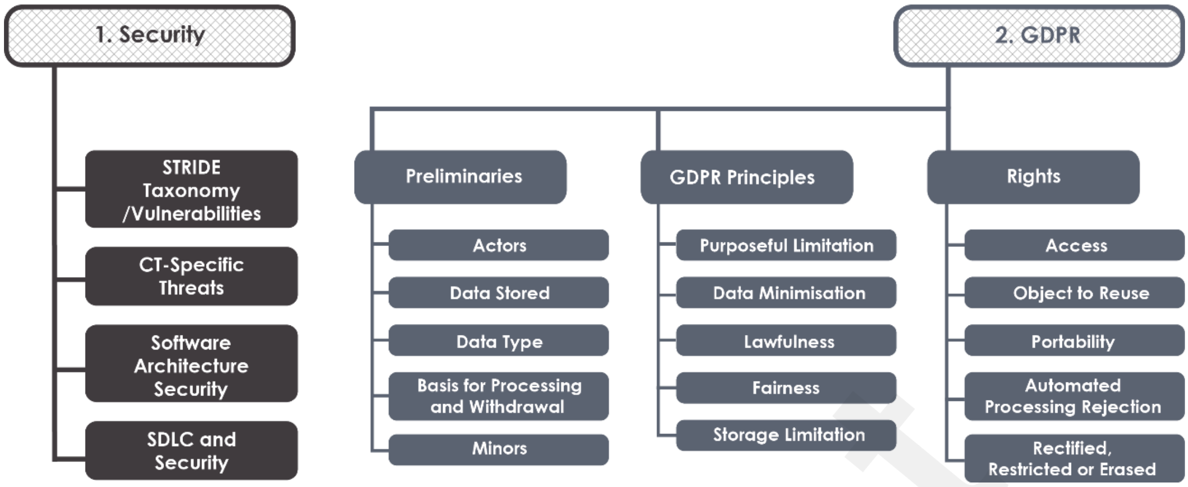


Figure 5. The Data Protection pillar.

Security consists of 4 criteria, which center around contract-tracing-specific security threats and vulnerabilities. They are scoped to ensure that CTAs are compared fairly, such that security vulnerabilities related to software or system components which cannot be changed by the CTA development are not considered; for example, those related to the system security of 3rd party providers (third-party vendors are noted in 'General Characteristics', under 'development partners', so an indirect warning flag is retained). These attributes incorporate a novel approach to CTA evaluation that was developed to ensure a lightweight comparison using the potentially incomplete data available for each app: analyzing vulnerabilities of distinct app functionalities against a common threat assessment model [82]. Attributes under Security (namely: STRIDE Taxonomy/Vulnerabilities, CT-Specific Threats, Software Architecture Security, and SDLC and Security) are designed to indicate whether these vulnerabilities are bugs in code, which can be fixed or would require a redesign of the architecture to address. For example, when we used this approach to compare the security of two CTAs (Corona-Warn [77] and MyTrace [83]), we first used automated tools to identify Common Weakness Enumerators (CWEs - a categorized 'encyclopedia' of over 600 types of software weaknesses), and then we manually confirmed them using in house security expertise. We compared the identified Enumerators for both apps (Table 2) and then against the predefined common threat assessment model (Table 3), providing an answer to our Security questions. Our analysis showed that while both apps may suffer from similar concerns related to information disclosure and de-anonymization, the CWEs which enable these are different, with our questions under SDLC and Security allowing us to see if the identified vulnerabilities can easily be patched/fixed or would be more difficult to correct.

Table 2. Comparison of Common Weakness Enumerators in Corona-Warn [77] and MyTrace [83].

CWE	Corona-Warn	MyTrace
89 - Improper Neutralization of Special Elements used in an Structured Query Language (SQL) Command	Local SQL injection possible but data encrypted	Local SQL injection possible and data not encrypted
276 - Incorrect Default Permissions	-	Permissions for tasks, Bluetooth admin, and external storage
295 - Improper Certificate Validation	Vulnerable to SSL MITM attack	-

532 - Insertion of Sensitive Information into Log File	Sensitive is Encrypted	Excessive information logged
327 - Use of a Broken or Risky Cryptographic Algorithm	Weak hash function in Secure Socket Layer (SSL)	-

Table 3. Comparison of Threats to Corona-Warn [77] and MyTrace [83] using the CWE's listed in Table 2 (with a severity rating H=High, M=Medium and L=Low) against the common threat assessment model.

Threat	Corona-Warn Matched CWEs	MyTrace Matched CWEs
Fake Alert Injection	-	CWE-327 H
False Report	CWE-295-H CWE-327-L	-
Proximity beacons altered	-	CWE-89-H
User can deny or retract infection report or contact details	CWE-295-H	-
Personal information disclosed	CWE-327-L CWE295-H	CWE-89-H CWE-276-H
User deanonymized and tracked	CWE327-L CWE295-H	CWE-89-H CWE-276-H CWE-532-H
Energy Resource Drain Attack	-	CWE-276-H
System Resource contention	-	CWE276-H

GDPR considerations are important because they talk to the essential user concern of data privacy. The GDPR attribute has three sub-attributes including Preliminaries, GDPR Principles, and Rights. Preliminaries involves information required for the evaluation of the individual data stored later: Data Stored, Data Type, and Basis for Processing. It also includes Withdrawal and whether the organization declared consent, has a legal requirement as their basis for processing the data. For instance, applying question DP06 (Data Stored): 'What personal data is collected?' to the COVID Tracker App [74] will generate the following list: phone number, date of last exposure, sex, age range, county, town, symptoms, diagnosis keys, date of symptom onset, app metrics, IP address, and app security tokens. GDPR Principles refer to the key principles of GDPR (such as Minimization, Fairness, and Storage Limitation), which are not under scrutiny in other dimensions of our C^{3}EF framework. It consists of 5 attributes which are evaluated across the stored data retrieved from the question on Data Stored. Most of these criteria require detail from the data controller and must rely on those details being truthful and accurate. Our final attribute under GDPR is Rights which refer to the rights of the individual that must be upheld if they are the subject of data processing by an organization. Therefore, they refer to the availability of organizational procedures to ensure these rights. We have 5 attributes under GDPR Rights: Access, Object to Reuse, Portability, Automated Processing Rejection, and Rectified, Restricted or Erased (data). An interesting counter-example here, is the AMAN app [78], which is not based in an EU state and, as such, is only required to adhere to the GDPR if it is used by EU citizens. It, therefore, had a distinct lack of documentation to support GDPR rights, as is to be expected.

Effectiveness

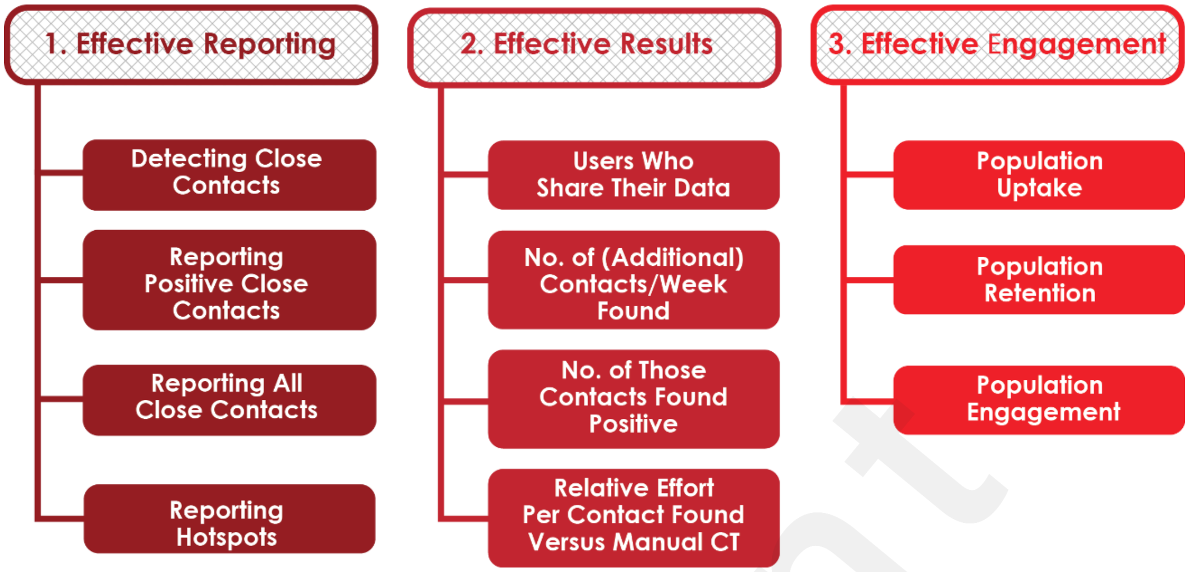


Figure 6. The Effectiveness Pillar.

The Effectiveness pillar refers to the degree to which the app is successful in its core aims: accurately detecting close contacts and providing relevant information to both citizens and health authorities. It contains three high-level attributes reflecting these concerns: *Effective Reporting*, *Effective Results*, and *Effective Engagement*, and consists of 11 attribute/questions combinations in total (see Figure 6).

Effective Reporting focuses on the ability of the app to report accurately on close contacts and the location of virus hotspots to individual users. It first assesses the accuracy of close contact detection: often, this will have to be reported at protocol level, for example, stating that the app reports at GAEN-level accuracy [66]. The framework then breaks down the reporting of close contacts into two categories: reporting on COVID-positive close contacts and reporting on a user's total number of close contacts over time periods. This latter category is sometimes reported in an effort to highlight and refine users' behavior, as in the case of PathCheck SafePlaces [75]. Finally, several of the apps report on prevalence information by locale, to let users know areas where the virus is more (or less) prevalent. This is a form of hotspot identification for users (see Figure 7 for an example from the HSE's COVID Tracker app [74]). The final question in this section probes the availability/granularity of this hotspot-location facility (Electoral division in the case of Figure 7).

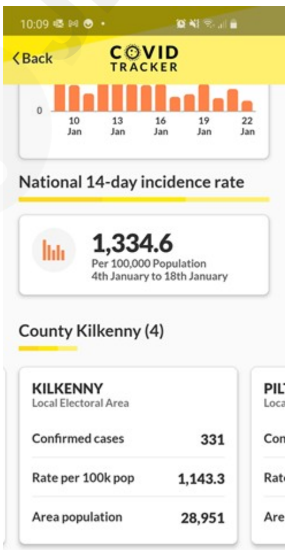


Figure 7. Electoral-District breakdown of national figures on the HSE COVID Tracker app [74].

Effective Results focuses on the ability of the app to meet its wider objectives across the jurisdiction. Hence, it concentrates on the (total) number of users who choose to share their data after being told of their positive status and the number of additional close contacts that are informed, based on that sharing. It then looks at the number of those contacts who were subsequently identified as positive. Finally, it aims to assess the relative time and effort in identifying a close contact via the app as opposed to via the manual-tracing effort.

Effective Engagement focuses on the population's adoption of the app, probing the uptake of the app across the population, citizens' retention of the app over time and, in cases where the app contains interactive features, the population's engagement with the app, as possibly measured by their usage of these interactive features. This is important because digital contact tracing is very dependent on the proportion of the population who upload the associated app and retain it over time.

Transparency

Transparency is discussed here as an independent pillar, despite the obvious overlap with GDPR. This is because, here it addresses the transparency of the processes and artifacts utilized/formed during development of the app specifically. In this context, transparency is an important aspect from the perspective of the adopting citizen's confidence. The pillar has been divided into categories looking at *App Transparency, User Participation*, and *Data Transparency* (Figure 8).

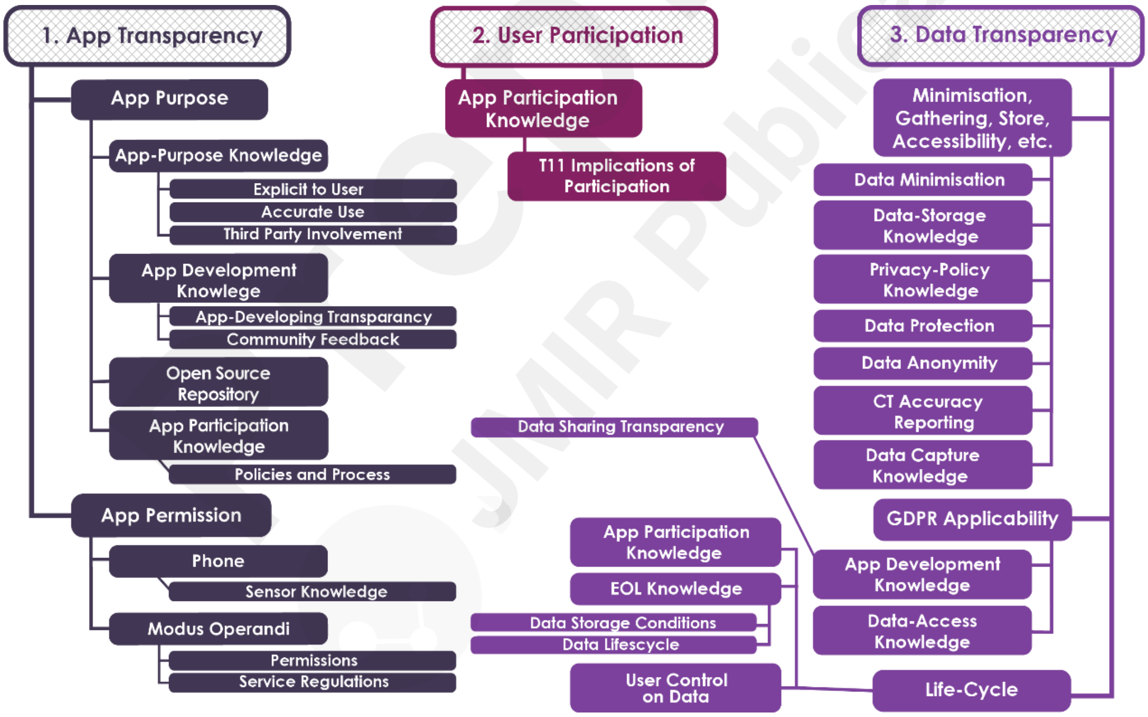


Figure 8. The Transparency pillar.

App Transparency includes App Purpose and App Permission. App Purpose offers the attributes: App-Purpose Knowledge, which refers to the purpose of the app being made accurately and accessibly explicit to the adopting citizen; App Participation Knowledge which looks at whether the citizens receive a clear explanation of the voluntary nature of participation; App Development Knowledge which looks at the mechanisms employed to guarantee community feedback; and; Open Source Repository to assess if the source code is made available (e.g. on GitHub), as this too shows transparency at a high-level. App Permission investigates all the permissions that are being asked for by the app, like permission to access Bluetooth/the camera, in terms of how transparent the app is about these phone functionalities accessed. Modus Operandi probes the CTA's transparency regarding the permissions required for its functionality, and looks at the time-period over which the services is being used as well as the contact tracing accuracy claimed by the developers (e.g. with questions like: 'Is the app being transparent about the contact tracing accuracy that they are achieving?').

User Participation consists of just one question, ensuring user consent: 'Does the app indicate and explain to the end-user about the voluntary nature of participation?'

Data Transparency focuses on whether the app has been designed following a Privacy-by-Design principle (under Data Capture Knowledge), as this will heighten confidence as to its data privacy nature: It assesses if the users are made accurately and accessibly aware of the data accessible to other bodies, both in terms of the data, and the accessing bodies (Data-Access Knowledge). It probes if the users are made explicitly aware of where and for how long their data is stored. 'Privacy policy knowledge' is also of concern here, looking at if, how, and when the citizen is informed about the data being collected in a Data Protection Impact Assessment. Two more attributes look at the 'minimality' of the collected personal data and at Data Protection, which focuses exclusively on the transparency section of data encryption and data anonymity. It has several questions which assess the anonymity, then encryption protocol, and the end-of-life conditions for the data.

Technical Performance

Technical Performance defines how efficiently a software system operates (in contrast to effectiveness), and it includes attributes and questions that help to capture this operational efficiency. Figure 9 shows the main attributes populating the Technical Performance pillar: Speed, Efficiency, Consumption, and Resource and Troubleshooting & Trust.

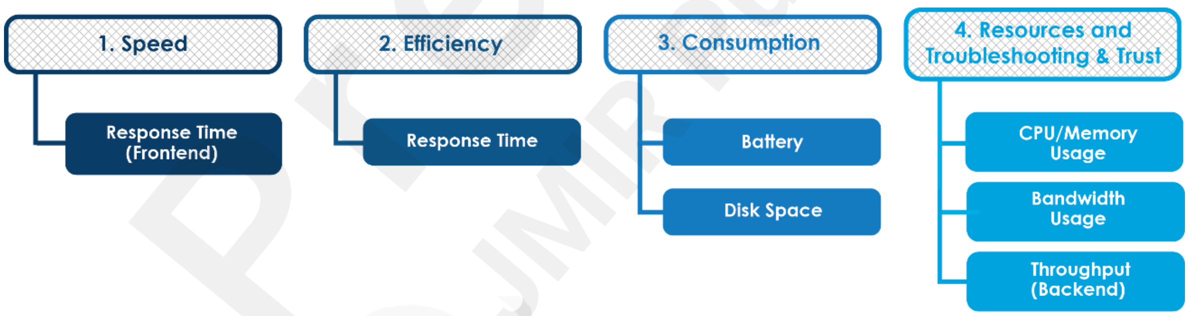


Figure 9. The Technical Performance pillar.

The ***Speed*** sub-attribute captures how quickly a software system's front-end application responds to a user's requests as delays may cause user frustration. This sub-attribute probes two issues. The first queries how fast the application responds to a user's interaction. The response here is measured in time units (for example, milliseconds), and can be influenced by several aspects, such as third-party applications (including an operating systems' libraries and API), hardware and its configurations, various components of the application, and how they work together.

The second ***Efficiency*** question focuses on the algorithms of the application that are responsible for answering a user's requests.

Consumption, Resources and Troubleshooting & Trust capture how efficiently a software system consumes available hardware resources – efficiency of battery usage, disk usage, CPU and memory usage, and bandwidth consumption. These attributes are particularly important with respect to retention if users perceive a battery/storage/CPU drain on their device [67] and so should probably be assessed as above or below some sort of 'noticeable-threshold' level.

Citizen Autonomy

The Citizen Autonomy pillar refers to the degree to which a user has the ability to define their own control levels in terms of the rights and accesses they grant the app. Additionally, it concerns itself with the user's ability to influence the evolution of the app going forward: an important element of autonomy, given that jurisdictions are asking users to retain the app for the duration of the emergency.

The pillar has three high-level attributes: App Discussion Authority, Phone Functionality, and Data Control (see Figure 10). Cumulatively these three categories consist of 9 questions.

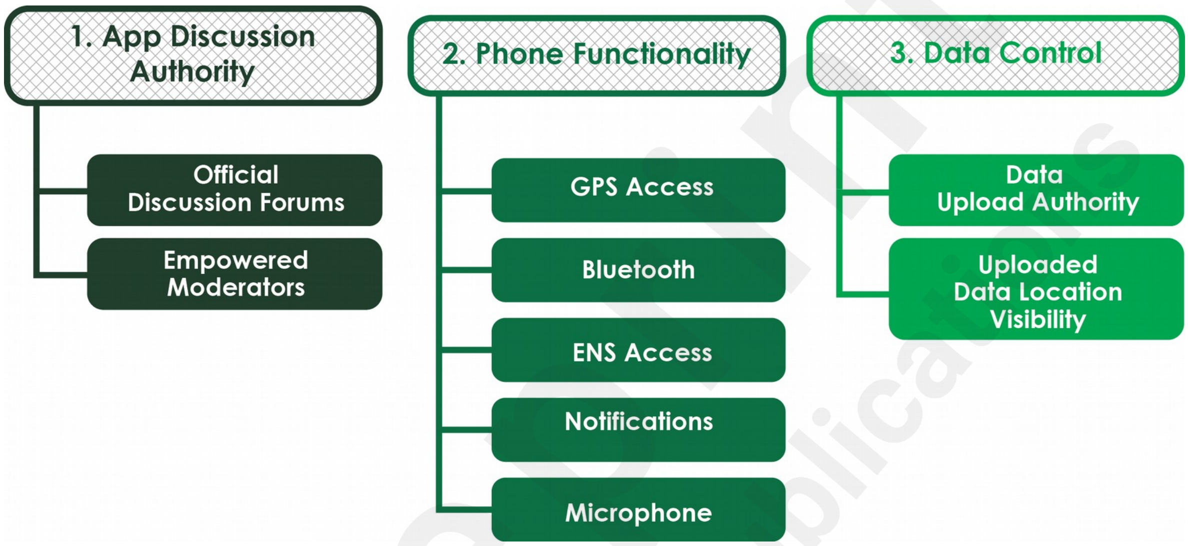


Figure 10. The Citizen Autonomy pillar.

App Discussion Authority focuses on the ability of the user to influence the future direction of the app and thus feel a sense of ownership. It first checks if there is a discussion forum where users are free to leave opinions and requests for change (most apps have at least a review section on Google Play or Apple's Appstore by default). An important consideration then is whether the available review sites are curated or moderated by representatives of the App development team. For example, the HSE's COVID Tracker app [74] has reviews on Google Play and the Apple Store, but it also has a GitHub repository [84] where users can leave their push requests for developers, as illustrated in Figure 11.

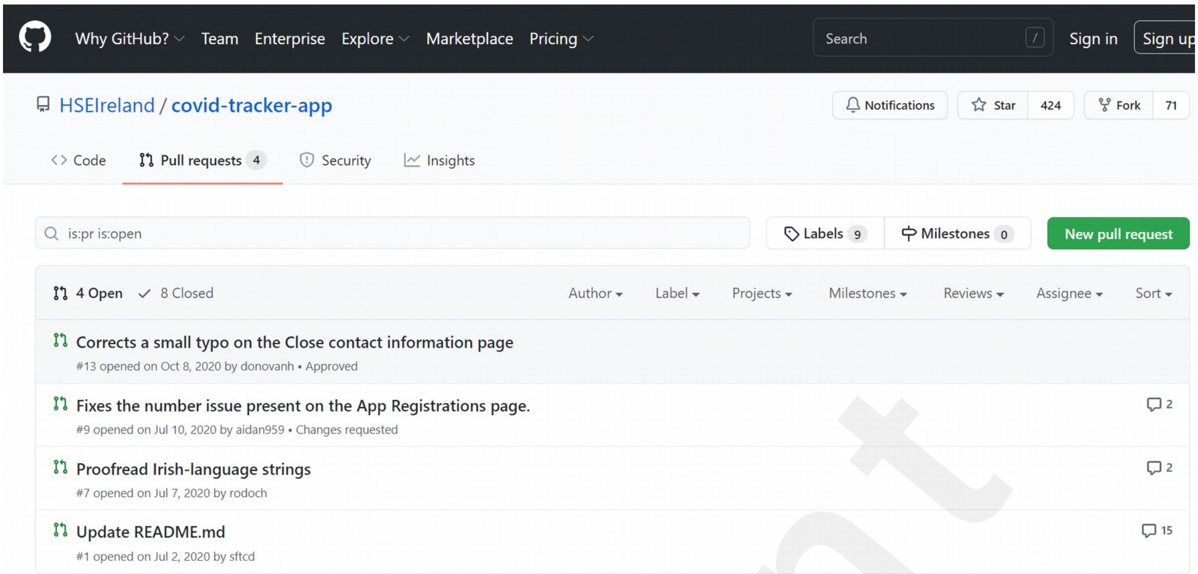


Figure 11. The ‘pull requests’, GitHub page [84] for the HSE’s COVID Tracker app [74].

Phone Functionality focuses on the ability of the user to control the app’s access to phone services. Typically, these are relevant services like GPS, GAEN [66], Bluetooth, and notifications, where their role in the app’s functioning is apparent. But occasionally, additional services may be required. For example, NOVID [76] uses ultrasound in an attempt to make close-contact detection more accurate and, as a result, requires access to the phone’s microphone.

Data Control focuses primarily on the data that is uploaded from the user’s app, typically in an instance of a positive diagnosis for the virus. It checks if the user is explicitly asked their permission for the upload of this sensitive data to happen and if the user is made aware of where the resultant data is stored (and by whom).

Conclusion

In this paper, we have reported on the development of an evaluation-concern framework to compare and contrast CTAs. Specifically, we developed C^3EF for CTAs - a framework of 7 pillars and 161 attributes, with related questions (180), whose primary concern is citizen adoption. We have reported on the iterative methodology and offered an overview of our final results.

One important consideration at this stage concerns the open-ended nature of C^3EF. The mixed methodology using both top-down and bottom-up strategies made particularly apparent how CTAs are constantly changing and evolving. Not only are CTAs (more or less) frequently updated to fix issues and improve functionalities, but they also operate in a changing scenario, which offers new requirements and design opportunities over time. For instance, at the beginning of our project, vaccines were not available. But, at the time of writing, a number of vaccines are being administered, which opened up new potential needs (digital vaccine passports, interoperability between systems when traveling abroad) that might extend the scope of current CTAs as we know and evaluate them now. We do not see this as a limitation of our work, but rather as an invitation to progress to further research on this topic and extending the framework. Reporting on methodology, but also sharing a supporting website gives convenient access to stakeholders [80], who can consider not only adopting but also adapting and further evolving our work. It is our intention that stakeholders will consider the included evaluation concerns as guidance when designing, evolving or evaluating CTAs, and that researchers will criticize and improve on those attributes to support improved and expanded CTAs over time.

We note the tension between the importance of an attribute and the feasibility of getting the information to answer the attribute's question. A good example is presented in the description of Phase 2: Concretization and Critique, where the number of users tested early in response to an app notification is considered a core effectiveness measure. But this is difficult to ascertain without buy-in from the national health authority. In a sense, we believe that the framework should ultimately contain important attributes even when it is difficult to obtain an answer, as they often highlight the core criteria that should be used to assess the apps and thus reflect a need for wider buy-in by the associated health authority and/or citizens.

This feasibility issue also highlights a limitation of this work. The framework focuses on the app themselves, but in order to have a significant impact on the spread of the virus, CTAs need an ecosystem of organizational, political and social entities which goes beyond the scope of our framework. The examples discussed in the previous paragraph assumes, for instance, that data about the number of citizens alerted of their COVID-19 positivity and the number who decide to get tested because of a notification in a CTAs exist, and that these numbers are monitored. However, this is only possible if certain procedures exist and are put in place by health authorities (such as keeping a record of individuals who asks for a test because of a notification from their CTAs). In other words, scoring well on our C^3EF does not guarantee that the CTAs will have an evaluatable impact on containing the spread of the virus.
The development of the C^3EF framework offers opportunities to improve CTAs. Through the descriptive questions, relevant aspects can be identified. However, as they are also 'prescriptive' in the sense that they give a notion of what is missing, they provide interesting information on what improvements can be made to existing and future CTAs.

Acknowledgements

The work undertaken was supported, in part, by Science Foundation Ireland (SFI) grant 13/RC/2094 and the 'COVIGILANT' Science Foundation Ireland grant 20/COV/0133

Conflicts of Interest

None declared

Abbreviations

CTA: contact tracing application
COVID-19: Coronavirus disease 2019
C^3EF: Citizen-Focused Compare-and-Contrast Evaluation Framework
API: Application Programming Interface
DP-3T: Decentralized Privacy Preserving Proximity Tracing
HCI: Human-Computer Interaction

Multimedia Appendixes

Multimedia Appendix 1

A summary of all papers identified for the literature review.

Multimedia Appendix 2

Grey literature sourced to identify CTA Characteristics.

Multimedia Appendix 3

A snippet of the Cross-pillar/ambiguity analysis matrix.

Multimedia Appendix 4

Visualization of all pillars and questions.

References

1. Baxter S, Goyder E, Chambers D, Johnson M, Preston L, Booth A. Interventions to improve contact tracing for tuberculosis in specific groups and in wider populations: an evidence synthesis. Heal Serv Deliv Res [Internet] National Institute for Health Research; 2017 Jan;5(1):1–102. [doi: 10.3310/hsdr05010]
2. Swanson KC, Altare C, Wesseh CS, Nyenswah T, Ahmed T, Eyal N, Hamblion EL, Lessler J, Peters DH, Altmann M. Contact tracing performance during the Ebola epidemic in Liberia, 2014-2015. PLoS Negl Trop Dis Public Library of Science; 2018 Sep 1;12(9). PMID:30208032
3. Anglemyer A, Moore T, Parker L, Chambers T, Grady A, Chiu K, Parry M, Wilczynska M, Flemyng E, Bero L. Digital contact tracing technologies in epidemics: A rapid review. Saudi Med J. Saudi Arabian Armed Forces Hospital; 2020. p. 1028. [doi: 10.1002/14651858.CD013699]
4. Danquah LO, Hasham N, MacFarlane M, Conteh FE, Momoh F, Tedesco AA, Jambai A, Ross DA, Weiss HA. Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: A proof-of-concept study. BMC Infect Dis Springer Nature; 2019;19(1). PMID:31533659
5. Ming LC, Untong N, Aliudin NA, Osili N, Kifli N, Tan CS, Goh KW, Ng PW, Al-Worafi YM, Lee KS, Goh HP. Mobile health apps on COVID-19 launched in the early days of the pandemic: Content analysis and review. JMIR mHealth uHealth. JMIR Publications; 2020. PMID:32609622
6. eHealth Network. Mobile applications to support contact tracing in the EU's fight against COVID-19: Common EU Toolbox for Member States. 2020.
7. ICO. COVID-19 Contact tracing: data protection expectations on app development [Internet]. 2020. [cited 2020 Aug 16] Available from: https://ico.org.uk/media/for-organisations/documents/2617676/ico-contact-tracing-recommendations.pdf
8. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics [Internet] 2020;46:427–431. [doi: 10.1136/medethics-2020-106314]

9. World Health Organization. Digital tools for COVID-19 contact tracing. 2020;(June).

10. Jacob S, Lawarée J. The adoption of contact tracing applications of COVID-19 by European governments. Policy Des Pract [Internet] 2020;1–15. [doi: 10.1080/25741292.2020.1850404]

11. Walrave M, Waeterloos C, Ponnet K. Adoption of a Contact Tracing App for Containing COVID-19: A Health Belief Model Approach. JMIR Public Heal Surveill 2020;6(3):e20572. [doi: 10.2196/20572]

12. Martin T, Karopoulos G, Hernández-Ramos JL, Kambourakis G, Nai Fovino I. Demystifying COVID-19 Digital Contact Tracing: A Survey on Frameworks and Mobile Apps. Wirel Commun Mob Comput Hindawi Limited; 2020;2020. [doi: 10.1155/2020/8851429]

13. Vinuesa R, Theodorou A, Battaglini M, Dignum V. A socio-technical framework for digital contact tracing. Results Eng Elsevier BV; 2020 Aug 1;8:100163. [doi: 10.1016/j.rineng.2020.100163]

14. Cho H, Ippolito D, Yu YW. Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs. arXiv Prepr arXiv200311511 [Internet] 2020; Available from: http://arxiv.org/abs/2003.11511

15. De Carli A, Franco M, Gassmann A, Killer C, Rodrigues B, Scheid E, Schoenbaechler D, Stiller B. WeTrace -- A Privacy-preserving Mobile COVID-19 Tracing Approach and Application. 2020;1–15. Available from: http://arxiv.org/abs/2004.08812

16. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: Framing the ethical challenges and how to address them. 2020 Apr 21; Available from: http://arxiv.org/abs/2004.10236

17. Dar AB, Lone H, Zahoor S, Amin Khan A, Naaz R. Applicability of Mobile Contact Tracing in Fighting Pandemic (COVID-19): Issues, Challenges and Solutions. 2020.

18. Bente BE, Van ’t Klooster JWJR, Schreijer MA, Berkemeier L, Van Gend JE, Slijkhuis PJH, Kelders SM, Van Gemert-Pijnen JEWC. The Dutch COVID-19 Contact Tracing App (the CoronaMelder): Usability Study. JMIR Form Res [Internet] JMIR Publications Inc.; 2021 Mar 1;5(3):e27882. [doi: 10.2196/27882]

19. Weiß JP, Esdar M, Hübner U. Analyzing the essential attributes of nationally issued COVID-19 contact tracing apps: Open-source intelligence approach and content analysis. JMIR mHealth uHealth 2021;9(3). PMID:33724920

20. ETSI. GR E4P 002 - V1.1.1 - Europe for Privacy-Preserving Pandemic Protection (E4P); Comparison of existing pandemic contact tracing systems [Internet]. 2021. Available from: https://www.etsi.org/deliver/etsi_gr/E4P/001_099/002/01.01.01_60/gr_E4P002v010101p.pdf

21. Wohlin C, Runeson P, Höst M, Ohlsson MC, Regnell B, Wesslén A. Experimentation in software engineering. Exp Softw Eng. Springer-Verlag Berlin Heidelberg; 2012. [doi: 10.1007/978-3-642-29044-2]ISBN:9783642290442

22. Lo B, Sim I. Ethical Framework for Assessing Manual and Digital Contact Tracing for COVID-19. Ann Intern Med 2020; [doi: 10.7326/m20-5834]

23. Lodders A, Paterson JM. Scrutinising COVIDSafe: Frameworks for evaluating digital contact tracing technologies. Altern Law J [Internet] 2020;45(3):153–161. [doi: 10.1177/1037969X20948262]

24. Sanwikarja P. Contact tracing: How do you design an app millions of people will trust? [Internet]. uxdesign.cc. 2020 [cited 2020 Jul 2]. Available from: https://uxdesign.cc/how-do-

you-design-an-app-millions-of-people-will-trust-8a63f5a5660a

25. Criddle C, Kelion L. Coronavirus contact-tracing: World split between two types of app - BBC News [Internet]. BBC.com. 2020 [cited 2020 Jun 23]. Available from: https://www.bbc.com/news/technology-52355028

26. Moss A, Spelliscy C, Borthwick J. Demonstrating 15 contact tracing and other tools built to mitigate the impact of COVID-19 | TechCrunch [Internet]. TechCrunch.com. 2020 [cited 2020 Jun 23]. Available from: https://techcrunch.com/2020/06/05/demonstrating-15-contact-tracing-and-other-tools-built-to-mitigate-the-impact-of-covid-19/?guccounter=1

27. Reuters. COVID-19 contact tracing apps: Here is how different countries are taking different approaches- Technology News, Firstpost [Internet]. Firstpost.com. 2020 [cited 2020 Jun 23]. Available from: https://www.firstpost.com/tech/news-analysis/covid-19-contact-tracing-apps-here-is-how-different-countries-are-taking-different-approaches-8369731.html

28. O'Neill PH, Ryan-Mosley T, Johnson B. A flood of coronavirus apps are tracking us. Now it's time to keep track of them. [Internet]. technologyreview.com. 2020. Available from: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/

29. Vokinger KN, Nittas V, Witt CM, Fabrikant SI, von Wyl V. Digital health and the COVID-19 epidemic: an assessment framework for apps from an epidemiological and legal perspective. Swiss Med Wkly NLM (Medline); 2020 May 4;150(1920):w20282. PMID:32418194

30. Morley J, Cowls J, Taddeo M, Floridi L. Ethical Guidelines for SARS-CoV-2 Digital Tracking and Tracing Systems. SSRN Electron J [Internet] 2020 [cited 2020 Jun 16]; [doi: 10.2139/ssrn.3582550]

31. Kasali FA, Taiwo OO, Akinyemi IO, Alaba OB, Awodele O, Kuyoro SO. An Enhanced Usability Model for Mobile Health Application. 2019;17(2):20–29. Available from: https://sites.google.com/site/ijcsis/

32. Kaur E, Haghighi PD. A context-aware usability model for mobile health applications. ACM Int Conf Proceeding Ser [Internet] New York, New York, USA: Association for Computing Machinery; 2016. p. 181–189. [doi: 10.1145/3007120.3007135]

33. Alonso-Ríos D, Vázquez-García A, Mosqueira-Rey E, Moret-Bonillo V. Usability: A critical analysis and a taxonomy. Int J Hum Comput Interact [Internet] Taylor & Francis Group; 2010 Jan;26(1):53–74. [doi: 10.1080/10447310903025552]

34. Soni N, Aloba A, Morga KS, Wisniewski PJ, Anthony L. A framework of Touchscreen interaction design recommendations for children (TIDRC): Characterizing the gap between research evidence and design practice. Proc 18th ACM Int Conf Interact Des Child IDC 2019 [Internet] Association for Computing Machinery, Inc; 2019. p. 419–431. [doi: 10.1145/3311927.3323149]

35. Gelman DL. Design for Kids: Digital Products for Playing and Learning. 2014. ISBN:9781933820309

36. General Data Protection Regulation (GDPR). Art. 8 GDPR - Conditions applicable to child's consent in relation to information society services [Internet]. GDPR.eu. 2018 [cited 2020 Oct 1]. Available from: https://gdpr-info.eu/art-8-gdpr/

37. EU. DIRECTIVE (EU) 2016/2102 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies. Off J Eur Union [Internet] 2016;327:1–15. Available from: https://eur-

lex.europa.eu/eli/dir/2016/2102/oj

38. ETSI. ETSI EN 301 549 - V2.1.2 - Accessibility requirements for ICT products and services. 2018 [cited 2020 Jul 17];2:1–152. Available from: https://www.etsi.org/deliver/etsi_en/301500_301599/301549/02.01.02_60/en_301549v020102p.pdf

39. w3cdevs. Web Content Accessibility Guidelines (WCAG) Overview | Web Accessibility Initiative (WAI) | W3C [Internet]. W3COrg. 2019 [cited 2021 Jul 18]. Available from: https://www.w3.org/WAI/standards-guidelines/wcag/

40. Silver A. Improving The Color Accessibility For Color-Blind Users [Internet]. Smashing Mag. 2016 [cited 2020 Aug 23]. Available from: https://www.smashingmagazine.com/2016/06/improving-color-accessibility-for-color-blind-users/

41. Xcertia. Xcertia mHealth App Guidelines [Internet]. Xcertia. 2019. [cited 2020 Aug 23]. Available from: https://www.himss.org/sites/hde/files/media/file/2020/04/17/xcertia-guidelines-2019-final.pdf

42. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth App Trustworthiness checklist. Digit Heal [Internet] SAGE Publications Inc.; 2019 Jan 21;5:205520761988646. [doi: 10.1177/2055207619886463]

43. Harrington CN, Ruzic L, Sanford JA. Universally Accessible mHealth Apps for Older Adults: Towards Increasing Adoption and Sustained Engagement. Int Conf Univers Access Human-Computer Interact Cham: Springer; 2017. p. 3–12. [doi: 10.1007/978-3-319-58700-4_1]

44. Kascak LR, Rebola CB, Sanford JA. Integrating Universal Design (UD) principles and mobile design guidelines to improve design of mobile health applications for older adults. Proc - 2014 IEEE Int Conf Healthc Informatics, ICHI 2014 Institute of Electrical and Electronics Engineers Inc.; 2014. p. 343–348. [doi: 10.1109/ICHI.2014.54]

45. Ballantyne M, Jha A, Jacobsen A, Scott Hawker J, El-Glaly YN. Study of accessibility guidelines of mobile applications. ACM Int Conf Proceeding Ser [Internet] New York, New York, USA: Association for Computing Machinery; 2018 [cited 2020 Jul 10]. p. 305–315. [doi: 10.1145/3282894.3282921]

46. Alturki R, Gay V. Usability Attributes for Mobile Applications: A Systematic Review. Cham; 2019.

47. Goel S, Nagpal R, Mehrotra D. Mobile Applications Usability Parameters: Taking an Insight View. In: Mishra D., Nayak M., Joshi A, editors. Inf Commun Technol Sustain Dev [Internet] Springer, Singapore; 2018 [cited 2020 Jul 5]. p. 35–43. [doi: 10.1007/978-981-10-3932-4_4]

48. Shitkova M, Holler J, Heide T, Clever N, Becker J. Towards Usability Guidelines for Mobile Websites and Applications. Wirtschaftsinformatik 2015;1603–1617.

49. Weichbroth P. Usability of mobile applications: A systematic literature study. IEEE Access Institute of Electrical and Electronics Engineers Inc.; 2020;8:55563–55577. [doi: 10.1109/ACCESS.2020.2981892]

50. Nielsen J. Usability engineering. Morgan Kaufmann; 1993. [doi: 10.1201/b16768]ISBN:9781439898567

51. Norman D. The design of everyday things: Revised and expanded edition. 2nd ed. Boulder: Basic Books; 2013.

52. Preece J, Sharp H, Rogers Y. Interaction design: beyond human-computer interaction. John Wiley & Sons. Chicago.; 2015.

53. Storni C, Tsvyatkova D, Richardson I, Buckley J, Abbas M, Beecham S, Chochlov M, Fitzgerald B, Glynn L, Johnson K, Laffey J, Mcnicholas B, Nuseibeh B, Connell JO, Keeffe DO, Keeffe IRO, Callaghan MO, Razzaq A, Rekanar K, Simpkin A, Walsh J, Welsh T. Toward a Compare and Contrast Framework for COVID-19 Contact Tracing Mobile Applications : a Look at Usability. In Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021) - Volume 5: HEALTHINF; 2021. p. 557–565. [doi: 10.5220/0010307005570565]

54. Ayala-Rivera V, Pasquale L. The grace period has ended: An approach to operationalize GDPR requirements. Proc - 2018 IEEE 26th Int Requir Eng Conf RE 2018 Institute of Electrical and Electronics Engineers Inc.; 2018. p. 136–146. [doi: 10.1109/RE.2018.00023]

55. European Commission. General Data Protection Regulation (GDPR) [Internet]. Gen Data Prot Regul. 2018 [cited 2020 Aug 25]. p. 1–99. Available from: https://gdpr-info.eu/

56. Data Protection Commission. Know Your Obligations [Internet]. [cited 2020 Jul 27]. Available from: https://www.dataprotection.ie/en/organisations/know-your-obligations

57. Information Commissioner's Office. Guide to the UK General Data Protection Regulation (UK GDPR) [Internet]. ICO; [cited 2020 Jul 27]. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

58. Horák M, Stupka V, Husák M. GDPR compliance in cybersecurity software: A case study of DPIA in information sharing platform. ACM Int Conf Proceeding Ser [Internet] New York, NY, USA: Association for Computing Machinery; 2019. p. 1–8. [doi: 10.1145/3339252.3340516]

59. Rios E, Iturbe E, Larrucea X, Rak M, Mallouli W, Dominiak J, Muntés V, Matthews P, Gonzalez L. Service level agreement-based GDPR compliance and security assurance in (multi)Cloud-based systems. IET Softw Institution of Engineering and Technology; 2019 Jun 1;13(3):213–222. [doi: 10.1049/iet-sen.2018.5293]

60. Torre D, Soltana G, Sabetzadeh M, Briand LC, Auffinger Y, Goes P. Using Models to Enable Compliance Checking against the GDPR: An Experience Report. Proc - 2019 ACM/IEEE 22nd Int Conf Model Driven Eng Lang Syst Model 2019 Institute of Electrical and Electronics Engineers Inc.; 2019. p. 1–11. [doi: 10.1109/MODELS.2019.00-20]

61. Robol M, Salnitri M, Giorgini P. Toward GDPR-compliant socio-technical systems: Modeling language and reasoning framework. Lect Notes Bus Inf Process [Internet] Springer Verlag; 2017. p. 236–250. [doi: 10.1007/978-3-319-70241-4_16]

62. Tom J, Sing E, Matulevičius R. Conceptual representation of the GDPR: Model and application directions. Lect Notes Bus Inf Process [Internet] Springer Verlag; 2018. p. 18–28. [doi: 10.1007/978-3-319-99951-7_2]

63. Allegue S, Rhahla M, Abdellatif T. Toward GDPR Compliance in IoT Systems. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) [Internet] Springer; 2020. p. 130–141. [doi: 10.1007/978-3-030-45989-5_11]

64. Tamburri DA. Design principles for the General Data Protection Regulation (GDPR): A formal concept analysis and its evaluation. Inf Syst Elsevier Ltd; 2020 Jul 1;91:101469. [doi: 10.1016/j.is.2019.101469]

65. Kammueller F. Formal Modeling and Analysis of Data Protection for GDPR Compliance of

IoT Healthcare Systems. Proc - 2018 IEEE Int Conf Syst Man, Cybern SMC 2018 Institute of Electrical and Electronics Engineers Inc.; 2019. p. 3319–3324. [doi: 10.1109/SMC.2018.00562]

66. Google. Exposure Notifications: Helping to fight COVID-19 – Google [Internet]. 2020 [cited 2020 Aug 20]. Available from: https://www.google.com/covid19/exposurenotifications/

67. Rekanar K, O'Keeffe IR, Buckley S, Abbas M, Beecham S, Chochlov M, Fitzgerald B, Glynn L, Johnson K, Laffey J, McNicholas B, Nuseibeh B, O'Connell J, O'Keeffe D, O'Callaghan M, Razzaq A, Richardson I, Simpkin A, Storni C, Tsvyatkova D, Walsh J, Welsh T, Buckley J. Sentiment analysis of user feedback on the HSE's Covid-19 contact tracing app. Ir J Med Sci [Internet] 2021 Feb 18;1–10. [doi: 10.1007/s11845-021-02529-y]

68. FOPH. SwissCovid - Apps on Google Play [Internet]. [cited 2020 Jul 20]. Available from: https://play.google.com/store/apps/details?id=ch.admin.bag.dp3t&hl=en

69. Ministry of Health and Centre for Disease Prevention and Control. Apturi Covid [Internet]. apturicovid.lv. 2020 [cited 2020 Jun 27]. Available from: https://apturicovid.lv/#en

70. Government of Italy. Immuni - Sito Ufficiale [Internet]. 2020 [cited 2020 Jul 20]. Available from: https://www.immuni.italia.it/

71. Troncoso C, Payer M, Hubaux J-P, Salathé M, Larus J, Bugnion E, Lueks W, Stadler T, Pyrgelis A, Antonioli D, Barman L, Chatel S, Paterson K, Capkun S, Basin D, Beutel J, Jackson D, Preneel B, Smart N, Singelee D, Abidin A, Gürses S, Veale M, Cremers C, Backes M, Binns R, Cattuto C, Persiano G, Fiore D, Barbosa M, Boneh D. GitHub - DP-3T/documents: Decentralized Privacy-Preserving Proximity Tracing -- Documents [Internet]. GitHub. 2020 [cited 2020 Aug 21]. Available from: https://github.com/DP-3T/documents

72. Decentralized Privacy-Preserving Proximity Tracing. 2020 [cited 2020 Jun 18];(April):33. Available from: https://en.wikipedia.org/wiki/Decentralized_Privacy-Preserving_Proximity_Tracing

73. COVID-19 apps - Wikipedia [Internet]. 2020 [cited 2020 Jun 18]. Available from: https://en.wikipedia.org/wiki/COVID-19_apps

74. HSE. Covid Tracker App - HSE.ie [Internet]. hse.ie. 2020 [cited 2020 Jun 30]. Available from: https://www.hse.ie/eng/services/news/newsfeatures/covid19-updates/covid-tracker-app/covid-tracker-app.html

75. Path Check Foundation. Home - COVID Safe Paths | COVID Safe Paths [Internet]. covidsafepaths.org. 2020 [cited 2020 Jun 27]. Available from: https://www.pathcheck.org/

76. novid.org. NOVID [Internet]. 2020 [cited 2020 Jun 16]. Available from: https://www.novid.org/

77. Open-Source Project Corona-Warn-App. Corona-Warn-App [Internet]. 2020 [cited 2020 Jun 29]. Available from: https://www.coronawarn.app/en/

78. Aman. Aman [Internet]. 2020 [cited 2020 Jun 7]. Available from: https://amanapp.jo/en

79. Overview ‹ Safe Paths — MIT Media Lab [Internet]. [cited 2020 Sep 25]. Available from: https://www.media.mit.edu/projects/safepaths/overview/

80. Digital Contact Tracing Taxonomy [Internet]. [cited 2021 Jan 30]. Available from: https://site-3663423-408-3579.mystrikingly.com/

81. Usability.Gov. User Interface Elements [Internet]. 2013 [cited 2020 Sep 20]. Available from:

https://www.usability.gov/how-to-and-tools/methods/user-interface-elements.html

82. Welsh T, Rekanar K, Abbas M, Chochlov M, Fitzgerald B, Glynn L, Johnson K, Laffey J, McNicholas B, Nuseibeh B, O'Connell J, O'Keeffe D, O'Keeffe IR, O'Callaghan M, Razzaq A, Richardson I, Simpkin A, Storni C, Tsvyatkova D, Walsh J, Buckley J. Towards a Taxonomy for Evaluating Societal Concerns of Contact Tracing Apps. 2020 7th Int Conf Behav Econ Socio-Cultural Comput Bournemouth, UK, 2020; 2020.

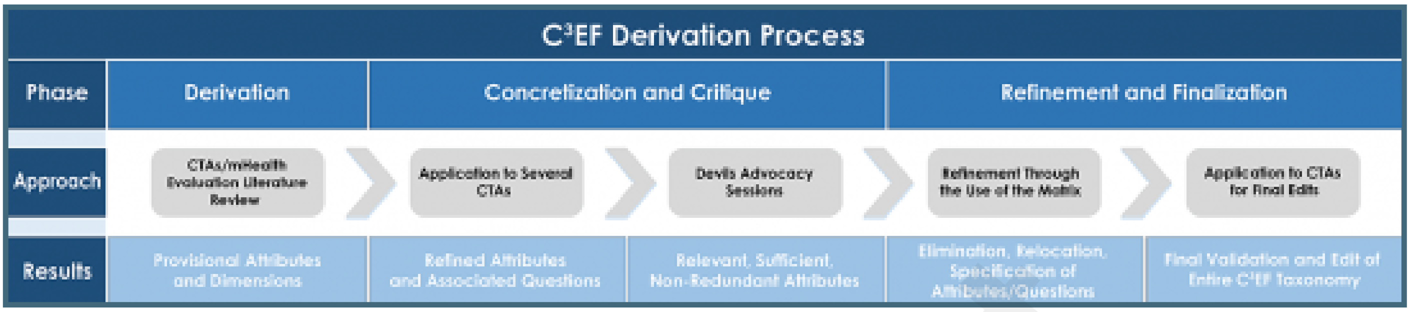
83. Ministry Of Science T and I (MOSTI). MyTrace, a Preventive Counter Measure and Contact Tracing Application for COVID-19 [Internet]. Off Portal Minist Sci Technol Innov. 2020 [cited 2020 Jul 25]. Available from: https://www.mosti.gov.my/web/en/mytrace/

84. HSE Ireland. HSEIreland/covid-tracker-app: GitHub [Internet]. [cited 2020 Jul 20]. Available from: https://github.com/HSEIreland/covid-tracker-app/pulls

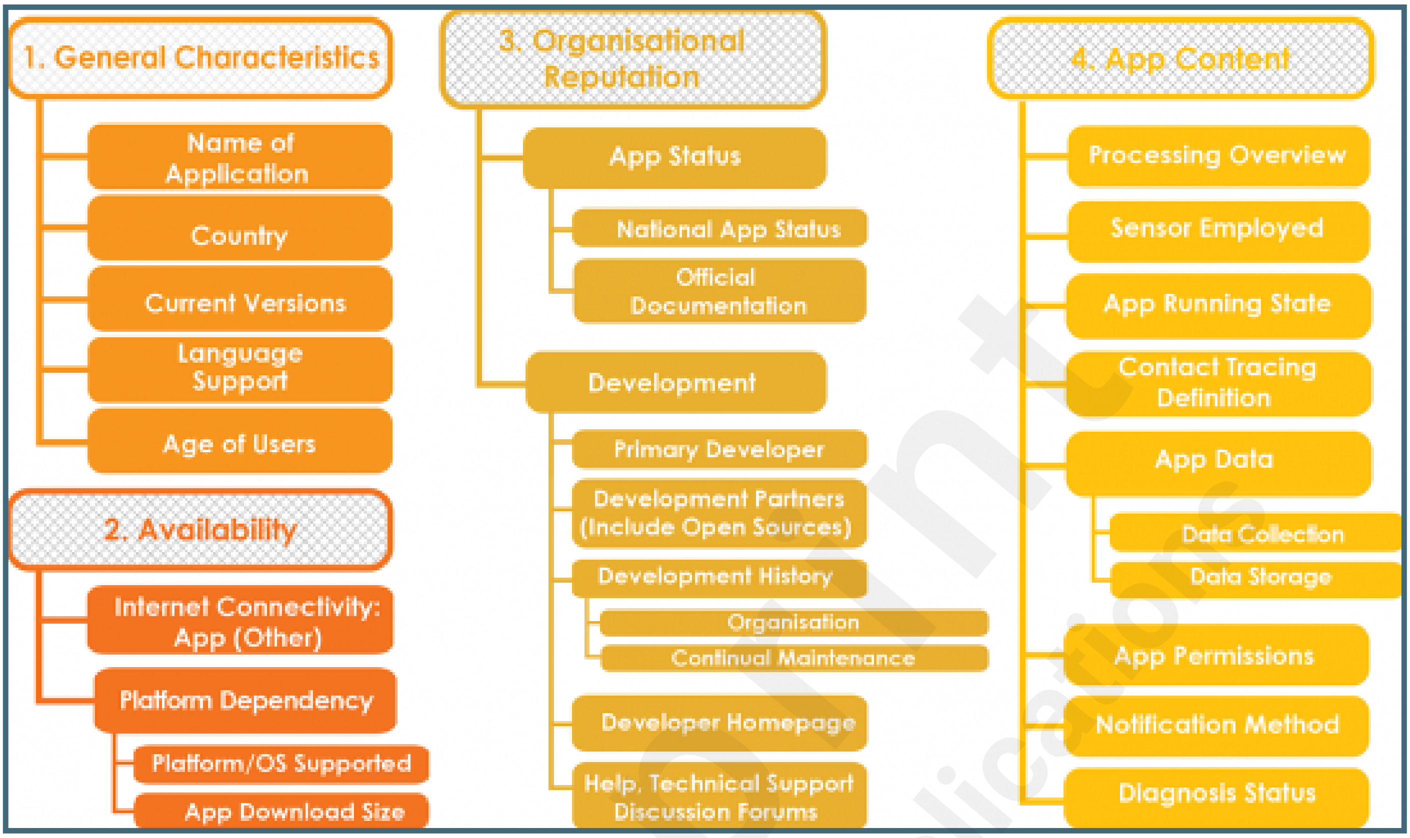
Supplementary Files

Figures

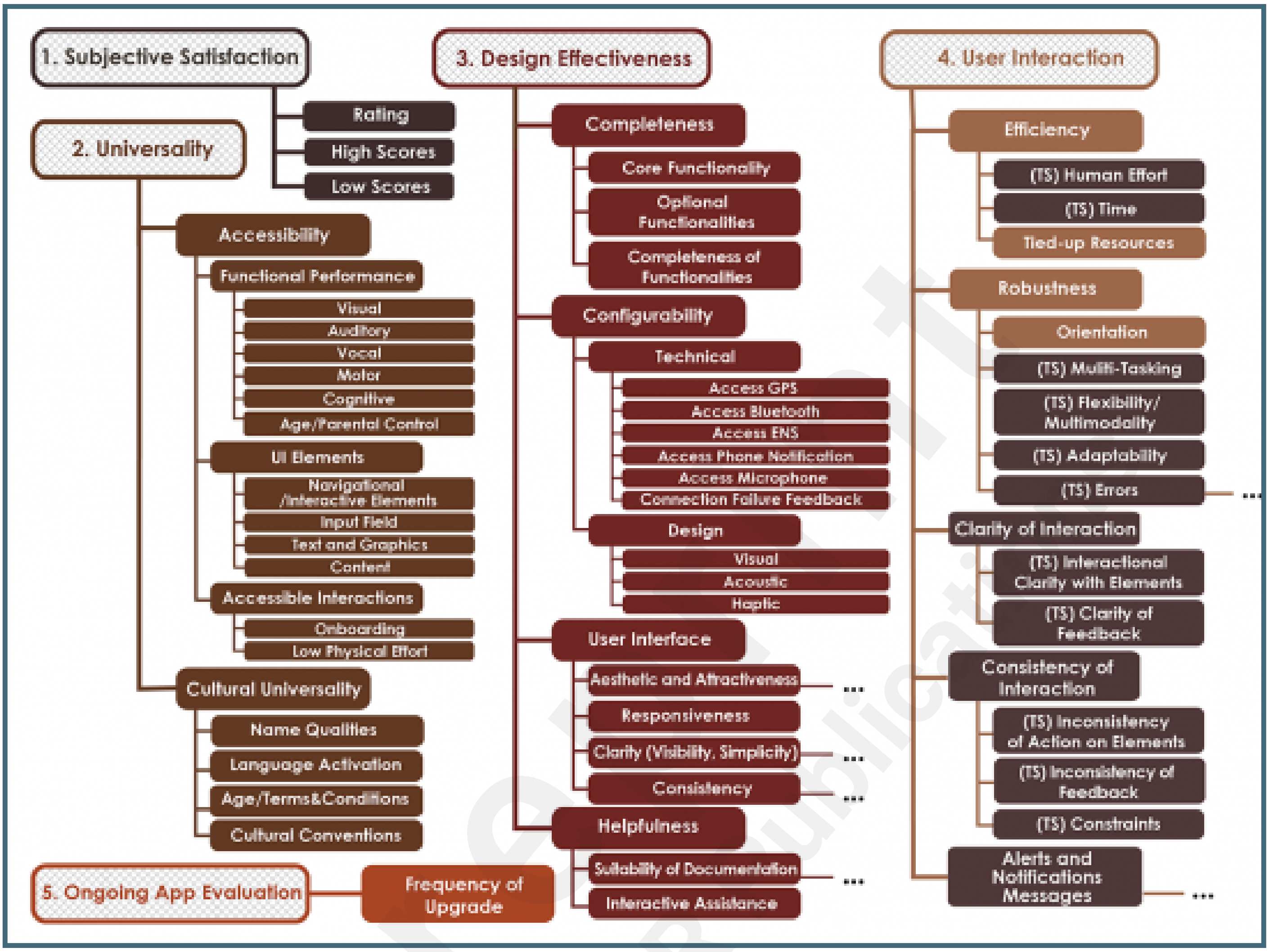
Activity phases and deliverables in the development of our Citizen-Focused Compare-and-Contrast Evaluation Framework for CTAs (C3EF) for CTAs.



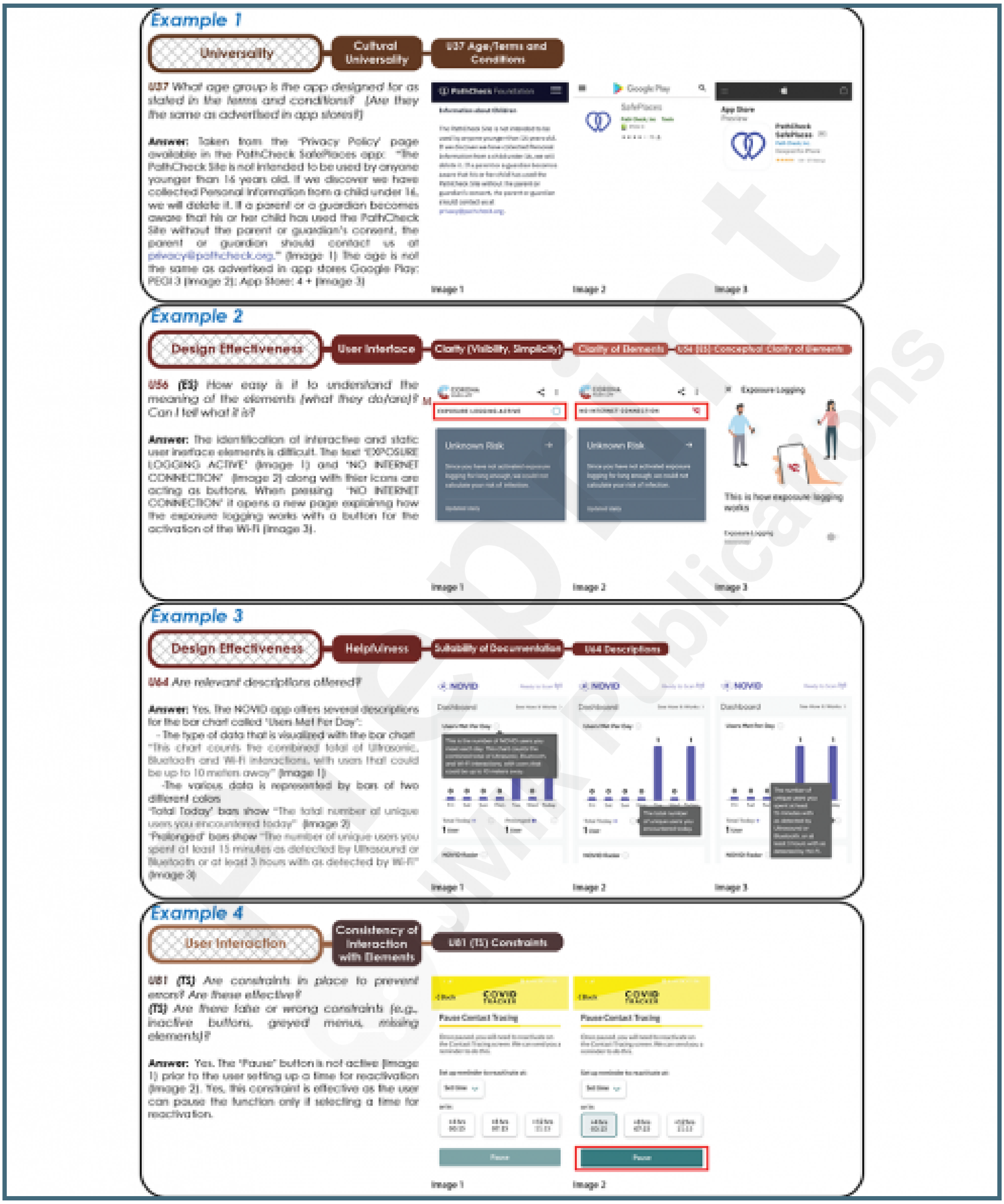
The Characteristics pillar.



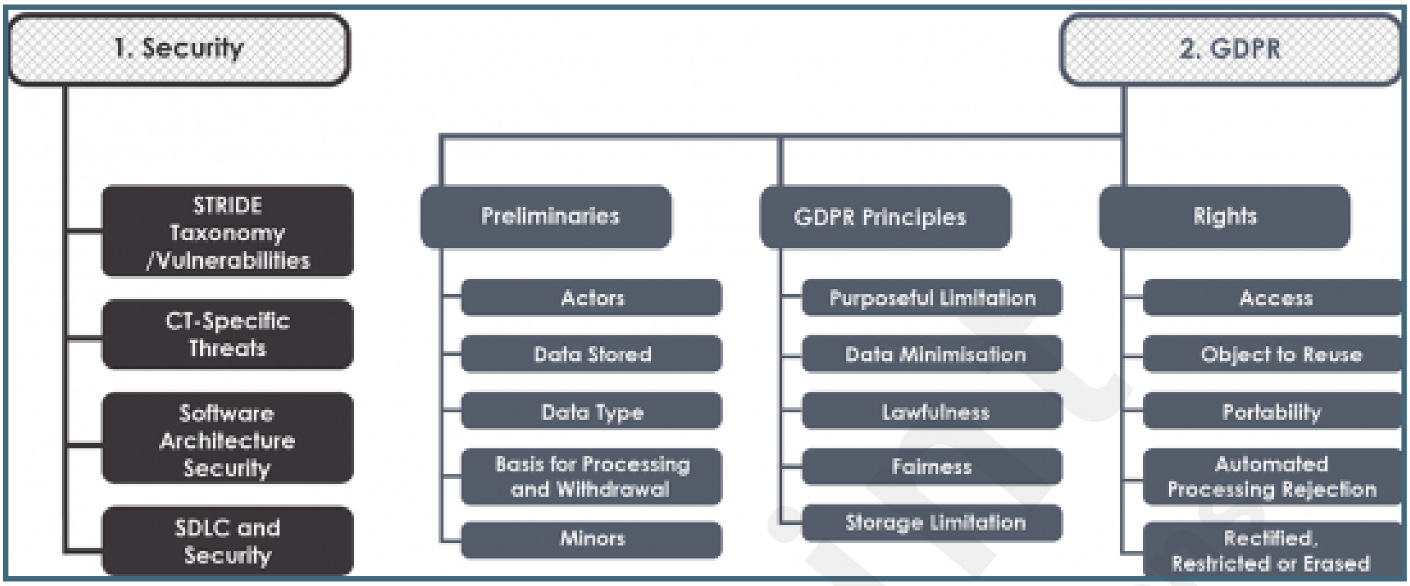
The Usability pillar.



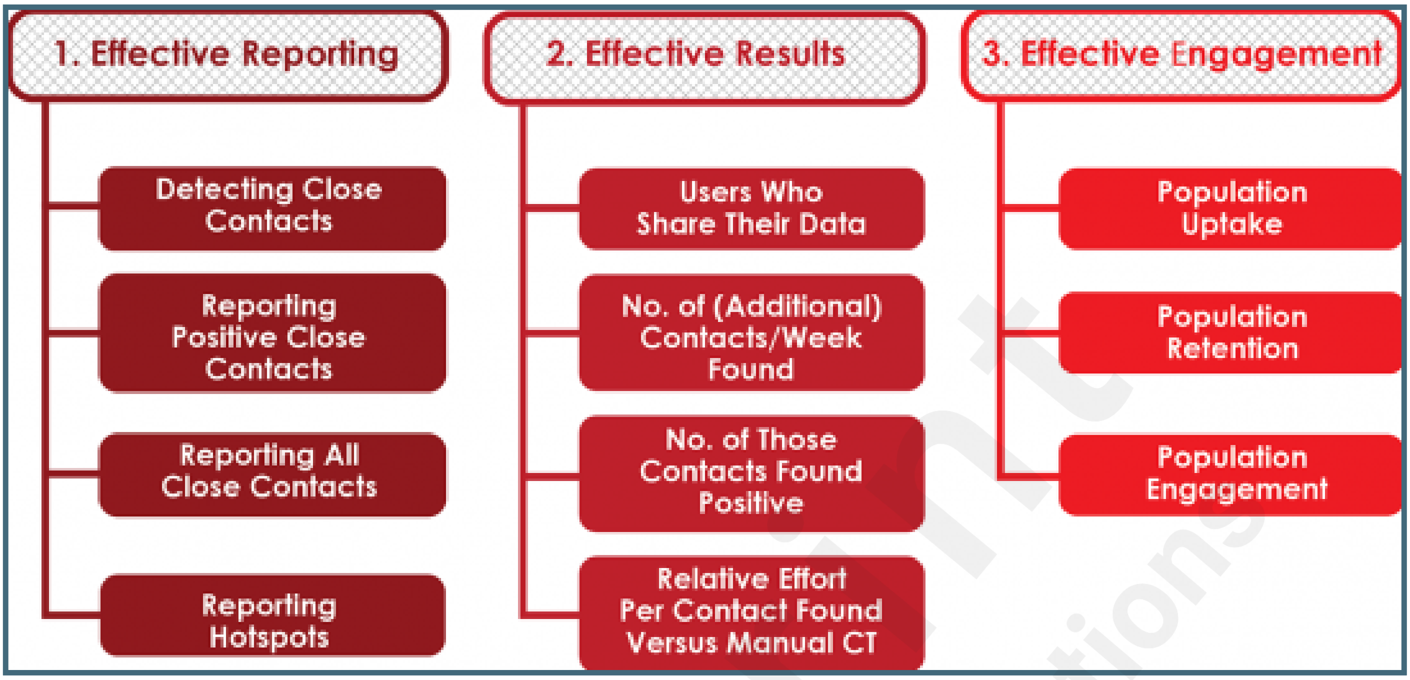
Example 1- PathCheck SafePlaces (US) [75]: age in ‘Terms of Use’. Example 2 - Corona-Warn (Germany) [77]: understandability of interface elements. Example 3 - NOVID app (US) [76]: description. Example 4 - COVID Tracker app (Ireland) [74]: constraints.



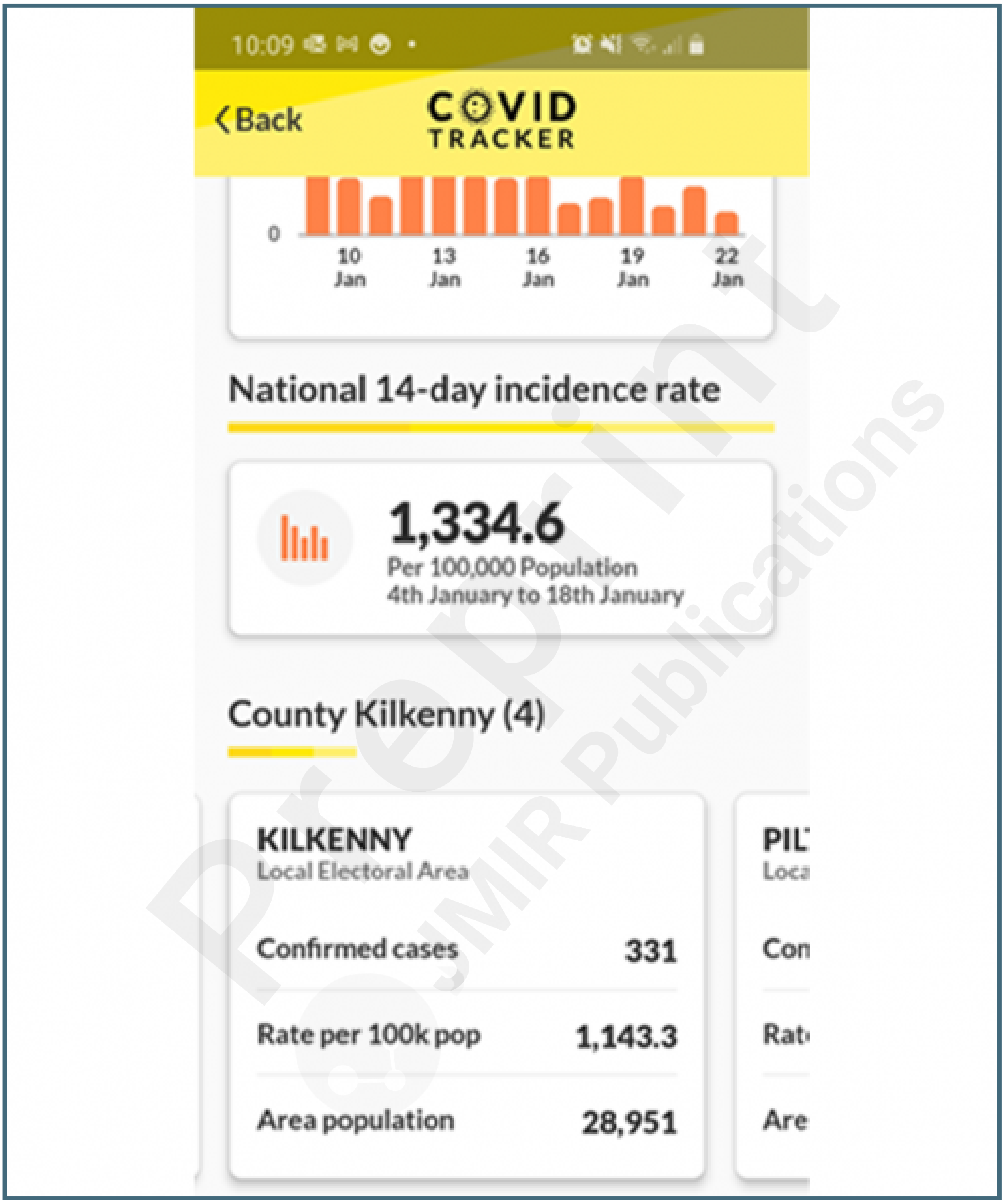
The Data Protection pillar.



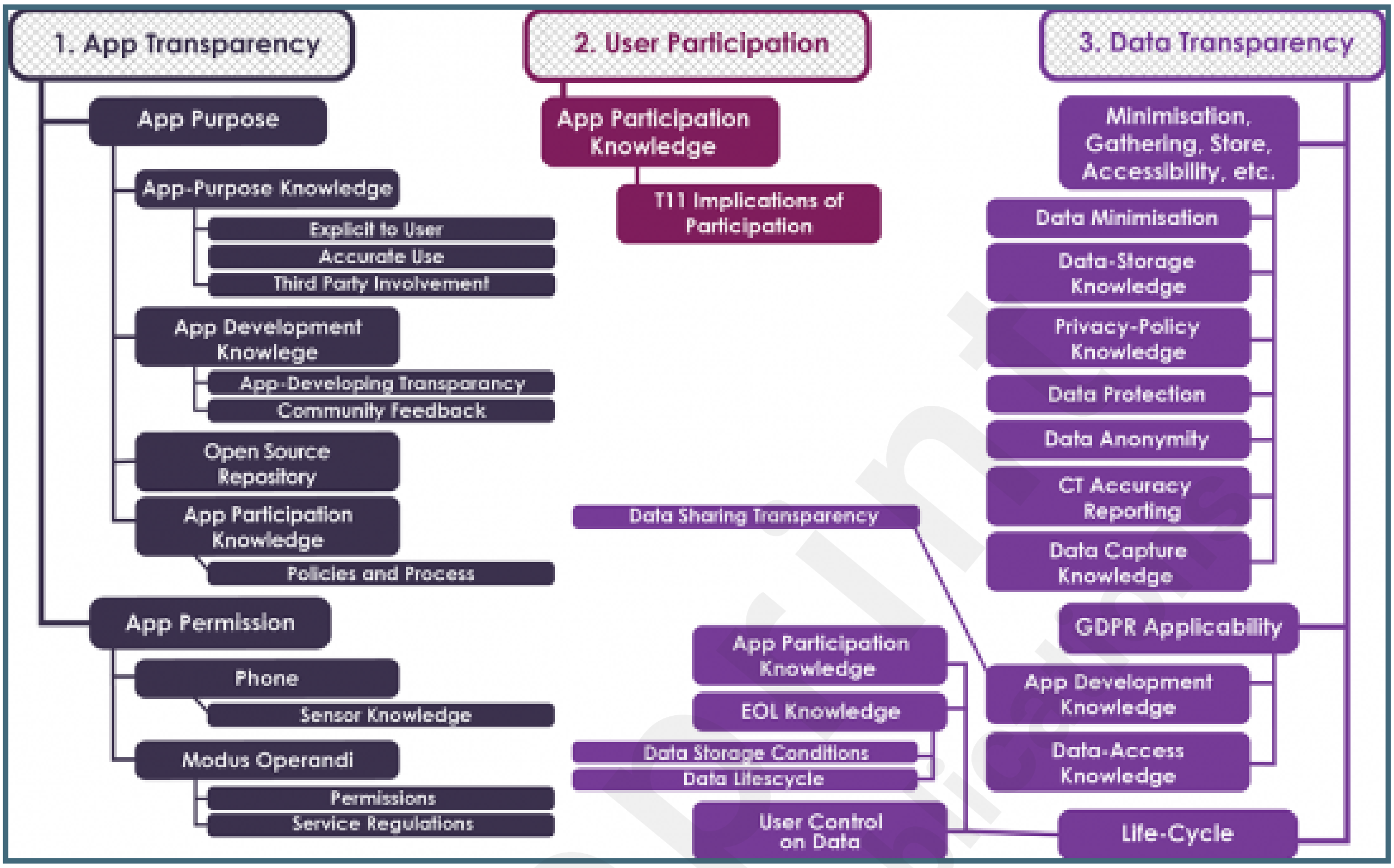
The Effectiveness Pillar.



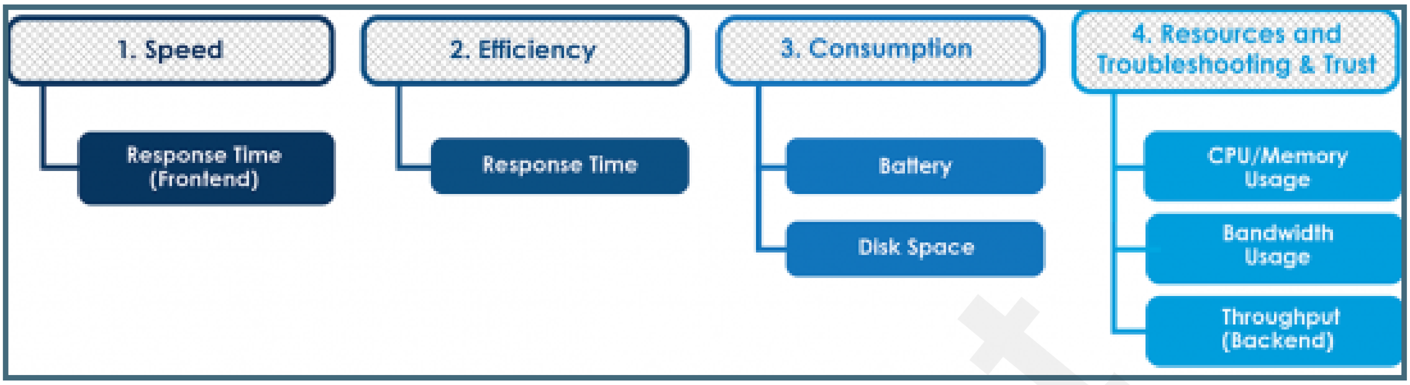
Electoral-District breakdown of national figures on the HSE COVID Tracker app [74].



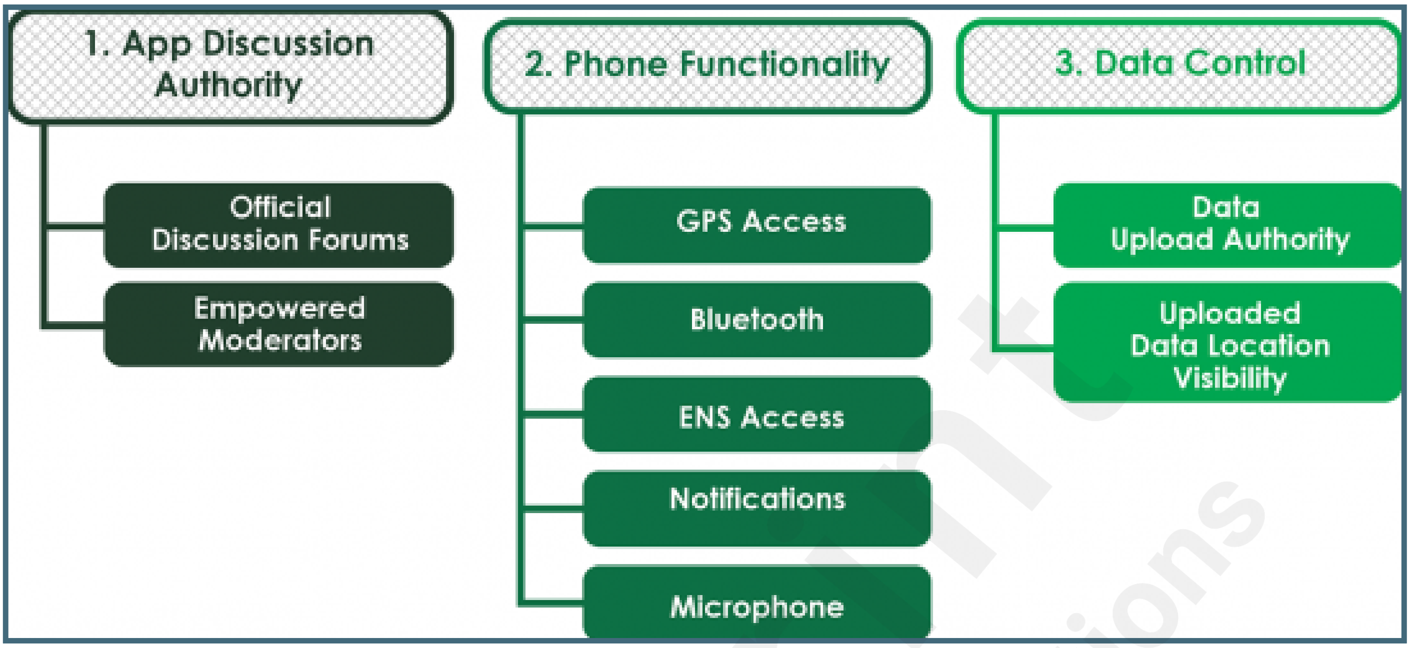
The Transparency pillar.



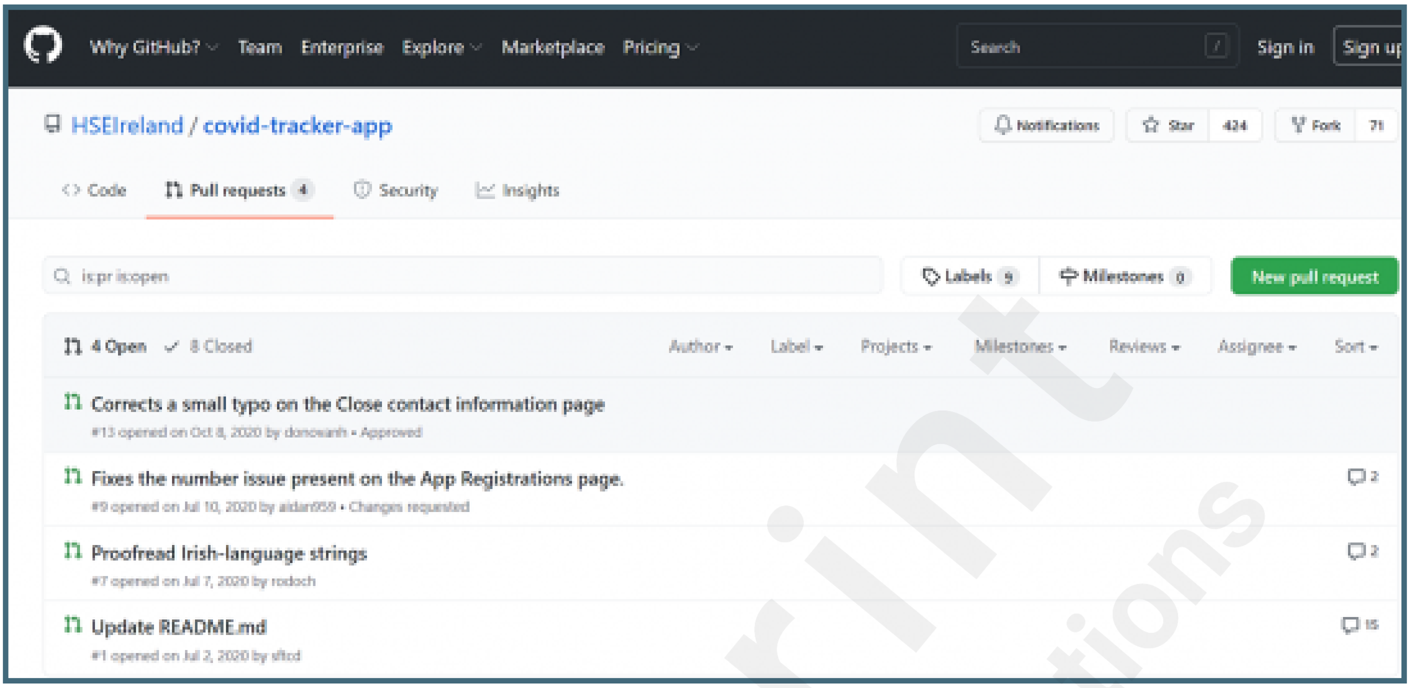
The Technical Performance pillar.



The Citizen Autonomy pillar.



The ‘pull requests’, GitHub page [84] for the HSE’s COVID Tracker app [74].



Multimedia Appendixes

A summary of all papers identified for the literature review.
URL: http://asset.jmir.pub/assets/aa00896d9228fdc844441bf9486f52db.docx

Grey literature sourced to identify CTA Characteristics.
URL: http://asset.jmir.pub/assets/d0ae69987a10f17a57b2b5fd0a91ba6d.docx

A snippet of the Cross-pillar/ambiguity analysis matrix.
URL: http://asset.jmir.pub/assets/c16d908f52936106dd0a0585d76e9deb.docx

Visualization of all pillars and questions.
URL: http://asset.jmir.pub/assets/16e875998db325810be40c8dfccbaed4.pdf