Towards Reproducible Indoor Positioning Research

Grigorios G. Anagnostopoulos[1], Alexandros Kalousis[1]

[1] *Geneva School of Business Administration, HES-SO, Geneva, Switzerland*

Abstract

The movement advocating for a more transparent and reproducible science has placed the issue of research reproducibility at the center of attention of various stakeholders related to academic research. Universities, funding institutions and publishers have started changing long established policies with the goal to encourage and support best practices for rigorous and transparent science making. Regarding the field of indoor positioning, there is a lack of standard evaluation procedures that would enable consistent comparisons. Moreover, the practices of Open Data and Open Source are on the verge of gaining popularity within the community of the field. This work, after presenting an extensive introduction to the landscape of research reproducibility and providing the viewpoint of the research community of Indoor Positioning, proceeds to its primary contribution: to provide a concrete set of suggestions that could accelerate the pace of the Indoor Positioning research community towards becoming a discipline of reproducible research.

Keywords

Reproducibility, Indoor Positioning, Open Science, Open Data, Open Source, Best Practices

1. Introduction

The concept of research reproducibility has gained increasing attention over the course of the last decade. Despite the many coordinated efforts that motivate the goal of more reproducible and transparent scientific results, many pitfalls are still identified. Indicatively, the lack of standardization in many disciplines, the unavailability of the data and the computer code which produced the published results and the '*insufficient peer review of published research*' combined with the absence of *'peer review of data*' [1], are some of the main pitfalls that have been underlined.

In its most general sense, the term *Reproducibility* of scientific results is often used as an umbrella term, covering a wide range of desirable attributes of science, including good quality, reliability and efficiency [1]. There is an ongoing effort of the scientific community to reach a consensus on the definition, in a strict and narrow sense, of the relevant terms laying below the overarching theme of Reproducibility (in its wider sense), such as: Reproducibility, Replicability and Repeatability. These terms are met in scientific publications with various, competing and even contradictory definitions [1,2]. There are even cases where the definitions assigned to the terms *Reproducibility* and *Replicability* are interchangeable across different works and scientific disciplines [3, 4].

In an effort to reach a consensus on the definitions of these terms, the National Academies of Sciences, Engineering and Medicine of the USA, have published a lengthy 'Consensus Study Report' on Reproducibility and Replicability in Science [2]. Upon extensive study of the current usage of the terms, '*the committee adopted definitions that are intended to apply across all fields of science and help untangle the complex issues associated with reproducibility and replicability*' [2].

The scoping Report of the European Commission on '*Reproducibility of scientific results in the EU*' [1], is in line with the American National Academies [2], in their definition of the terms of major interest: *Reproducibility* and *Replicability*.

IPIN 2021 WiP Proceedings, November 29 -- December 2, 2021, Lloret de Mar, Spain
EMAIL: {grigorios.anagnostopoulos, alexandros.kalousis}@hesge.ch
ORCID: 0000-0002-8643-7427 ; 0000-0001-6282-0686;

Reproducibility, in the strict sense, is achieved when new researchers are able to reproduce the analysis of the original authors, using the same '*input data, computational steps, methods, and code; and conditions of analysis*' [2], and obtain consistent results.

Replicability is achieved when new researchers are able to obtain '*consistent results across studies aimed at answering the same scientific question*' [2], '*using the same analytical method, but on different datasets*' [1].

In addition, the following terms are also commonly used:

Repeatability refers to the ability of the original authors of one work to repeat their experiment (or simply to run their code) under the exact same conditions (data, code, environment, methods, hardware) and obtain consistent results [5].

Runnability refers to the ability to obtain consistent results when executing the exact same steps on a new machine, using the same data, computational steps, methods, code and conditions of analysis [5].

Reusability refers to '*the looser possibility to re-use the results beyond the original research context, both inside and outside the original scientific discipline*' [1].

The current work presents an extensive overview of the landscape of research reproducibility, before providing the viewpoint of the research community of Indoor Positioning on this matter and proposing a list of concrete suggestions. In Section 2, the concepts of transparency, clarity and verifiability of evaluation are analyzed, with particular focus on Indoor Positioning research. The concept of Open Science, along with the various actions that it relates to, are discussed in Section 3. The ways that relevant best practices are being incentivized are extensively presented Section 4. Section 5 builds on the content of the previous sections to provide a concrete set of suggestions that could accelerate the pace of the Indoor Positioning research community towards becoming a truly Reproducible discipline. Lastly, conclusions drawn and future directions are discussed is Section 6.

2. Transparency, Clarity and Verifiability of Evaluation

Transparency, clarity, and verifiability of evaluation of scientific research are crucial values of science ethics. '*The integrity of datasets; the availability of data and the transparency of data collection methods (what was not reported, what was not used, why); the coherence of the approach (pre-registration of method/protocol)*' [1], are crucial elements that may enhance the reproducibility of scientific findings.

2.1. Verifiability of Evaluation in Indoor Positioning Research

The aspects related to the transparency and verifiability of evaluation, which are of great importance for all disciplines, are of particular interest when considering them in the context of Indoor Positioning research. The way data are collected, processed, and used for evaluation has not been standardized so far, by the community of the field. Nevertheless, the way these steps are performed greatly affects the outcome of the experiments utilizing them.

A particularly relevant work by Adler et al. [6] has studied the evaluation practices of Indoor Positioning research, performing a survey on papers from the IPIN conferences of the period 2010-2014 [6]. In their emblematic work, Adler et al. [6] analyzed 183 papers, categorizing them according to their *Ground Truth* collection method, their type of *Evaluation*, their type of *Reference System*, and their *Baselines*.

In terms of the *Evaluation* categories, Adler et al. [6] categorized works in different Evaluation approaches such as Discrete Point Experiments, Grid-like Experiments, and Office Walk Experiments. Undeniably, the *Reference Systems* used to establish the ground truth are very closely related to the evaluation process. Examples of Reference System categories are the use of Landmarks ('Single reference points with varying degree of accuracy'), of Paths ('e.g., fixed points on the floor or landmarks in the vicinity'), of Optical Systems (tracking the target using cameras) or of GNSS Reference systems.

Upon their analysis, the authors of [6] concluded that although '*most of all papers describe their setup very well, ... (they) tend to neglect the information on how the ground truth information was gathered.*' More particularly, regarding the ground truth definition, they observed a complete absence of reporting the way the time reference was obtained, in contrast to the commonly reported spatial reference. Various works have proposed protocols for the spatio-temporal definition of the ground truth and for its comparison to the obtained location estimates [7, 8]. The authors underline that following a well reported and rigorous methodology for the spatio-temporal definition of the ground truth is indispensable [6]. Lastly, Adler et al. [6] identified a systematic lack of external baselines in the evaluation section of the studied papers. With no baseline to compare against, it is hard to evaluate the potential of new methods and '*to quantify the progress made over the years*' [6].

The selection of the evaluation metrics is an important feature of complete reporting. It is often the case that single evaluation metrics are reported (mean, median, percentiles, standard deviation, Root Mean Square Error, Mean Square Error, etc.). In addition, error distributions depicted in the form of boxplots or of Cumulative Distribution Functions are also commonly reported, providing a better overview of the performance. Metrics that go beyond the Euclidean distance between the true and the estimated positions have recently been proposed. Such approaches take into consideration the particularities of the buildings, defining the positioning error as '*the length of the pedestrian path that connects the estimated position to the true position*' [9]. Since Indoor Positioning Systems (IPSs) may be used in different scenarios, facilitating very distinct services, there is no golden standard of an evaluation metric. It is therefore recommended for authors to provide the source code of their work, so that all relevant metrics can be easily calculated. Authors may choose to report the metrics relevant to their use case, but an open code approach would facilitate reusability, where multiple metrics may be evaluated.

An important aspect of transparent and verifiable evaluation is the availability of the used datasets. It is not sufficient to simply use a publicly available dataset, as the train/validation/test separation of the data should also be openly reported. Moreover, the principles of using the train/validation/test sets should be carefully respected. No evaluation should take place using the training set. Moreover, any potential tuning of hyperparameters should be performed on a validation set, which must be distinct from the (previously unseen by any other operation) test set on which results are reported. It is not uncommon to find works reporting performance on data that have been used either for training the model or for tuning certain hyperparameters, practices that enact the error of information leakage.

3. Open Science - Publicly Sharing Contributing Resources

The concept of Open Science describes a wide spectrum of actions that make '*the content and process of producing evidence and claims transparent and accessible to others*' [10]. The *Manifesto for reproducible Science* states that '*Transparency is a scientific ideal, and adding 'open' should therefore be redundant. Science often lacks openness: many published articles are not available to people without a personal or institutional subscription, and most data, materials and code supporting research outcomes are not made accessible, for example, in a public repository*' [10]. In this section we discuss a few main actions of openness, which have the potential to contribute to the goal of more reproducible research.

3.1. Open Data

A principal target of the movement of *Open Science* is the availability of *Open Research Data*, which can enhance reproducibility, verifiability and comparability of scientific results [11]. Sharing data can also facilitate the establishment of benchmarks and can save considerable time that would be required if all authors had to perform their own data collection.

A commonly used guideline for the way data should be shared is the *FAIR Guiding Principles* [12]. The *FAIR Principles* require data sharing in ways that guarantee that data are *Findable*, *Accessible*, *Interoperable*, and *Reusable*. *Findable* data should be assigned a unique and persistent identifier, should contain rich descriptive metadata and should be indexed in a searchable resource. *Accessible* (meta)data should be retrievable by their identifier using a standardized communication protocol that should be open, free, and universally implementable. Moreover, metadata should remain accessible, even when the data might no longer be available. *Interoperable* (meta)data should use a formal, accessible, shared, and broadly applicable language for knowledge representation, including qualified references to other (meta)data that may be relevant. *Reusable* data should be richly described with a plurality of accurate and relevant attributes, aiming at a clear and accessible data usage [12]. Including a license indicating the conditions under which the data can be used is of great importance. Various repositories support the sharing of open research data, such as the *Zenodo* repository or IEEE's *DataPort*.

3.2. Open Source

Open Source, or *Open Code*, is another very useful practice of *Open Science*, in which authors share the source code implementation that was used to produce the results of their published work. The code carries the potential to unambiguously present the exact way the reported results were produced. The shared code may concern all steps of the work, from the implementation of a new method that the paper may have proposed, to the experimental setting, the data digestion, and the results' calculation and their visualization. Simply sharing the code may not suffice for the code to be functional in other systems and to produce the same results. All package and library versions used, and all potential dependencies should be indicated. Authors could choose to share a file which can be used to recreate the environment used (such as the *yml* filetype). Lastly, it is crucial to share the random seed for non-deterministic operations.

Open Source should abide by the same standards as the ones discussed for *Open Data* [11]. More particularly, the code should be publicly available with a persistent, unique reference (like DOI). It should be well commented and should contain helpful metadata, explaining the content and guiding the users on the way the code can be used. Including a license, indicating the conditions under which the code can be used and extended is also very important.

3.3. CRediT – Contributor Roles Taxonomy

In the spirit of openness and transparency, the exact type of contribution of each researcher appearing in a paper's author list should be clearly stated. CRediT [13] is a widely used high-level taxonomy of Contributor Roles, representing the roles typically played by contributors to scientific scholarly output. Such public recognition of various roles of contributors can foster the collaboration of different teams and motivate data and material exchange. For instance, the pre-agreement on these publicly stated roles may remove the reservations of teams to share material and data with other teams, since the recognition of their contribution will have been agreed upon.

3.4. Registered Reports

Registered Reports is a publishing format in which authors submit for peer review a detailed description of a study/experiment they intend to undertake, describing the methodology and the protocol to be followed [14]. In Stage 1 review, fellow peers review a '*manuscript that includes an Introduction, Methods, and the results of any pilot experiments that motivate the research proposal*'. Upon the Stage 1 review, and its potential revisions, works with high quality protocols receive an *In-Principle Acceptance (IPA)*, which represents the promise that the final form of the paper will be accepted for publication if the authors follow through with the registered methodology, regardless of the outcome of the study [1]. The authors that receive the IPA can proceed to implementing the proposed experiment, collecting the data and enriching the initial manuscript with the sections presenting Results and Discussion. This complete version of the manuscript undergoes a Stage 2 review. After the potential revisions that may be suggested, the manuscripts are published. Naturally, authors are strongly encouraged to openly share data and code.

This publishing format can have numerous advantages. An important factor is that authors receive peer feedback on the process of designing their protocol, which can lead to the selection of more robust protocols. Moreover, this publishing format '*emphasizes the importance of the research question and the quality of methodology*' [1], rather than prioritizing positive results. At the time of writing this paper '*288 journals use the Registered Reports publishing format either as a regular submission option or as part of a single special issue*' [14].

4. Incentivizing Best Practices

The movement promoting Open Science and Research Reproducibility has inspired and motivated many researchers in refining their overall research workflow management with the aim of conforming to the highest standards of scientific rigor. Nevertheless, it is widely accepted that the facilitation of a widespread establishment of best practices requires an organized change and systemic support, as it cannot simply rely on the motivation of individuals. There has been a growing realization that '*little progress can be made if becoming involved in such activities reduces a researcher's chances of rank and status advancement and other rewards*' [15].

The fact that '*little emphasis is placed on the rigor of research when hiring, reviewing, and promoting researchers*' [16] has been identified as a major counter-factor, preventing the wide adoption of best practices. A related identified issue is that the '*novelty*' of research is systematically favored over '*rigour*' [16]. This phenomenon, often referred to as '*publication bias'* [17], is a strong feedback loop, fueling the incentive structure that favors the publication of novel results over the publication of negative results or of replication studies. The publication bias effect is accused of facilitating '*the dissemination and maintenance of false knowledge*' [17]. This is because, '*when incentives favor novelty over replication, false results persist in the literature unchallenged, reducing efficiency in knowledge accumulation*' [18].

A significant additional effort is required to produce scientific results that are reproducible, which raises the questions of how this additional cost is covered and how the adherence to these practices is rewarded. There is no easy way to accurately quantify this additional effort in a general manner. The stance of the relevant literature on estimating this cost varies from the rather optimistic view that '*authors can increase the trustworthiness and reproducibility of research results with relatively little effort*' [11], to the more pessimistic view, suggesting that '*researchers who adopt good practice in reproducibility are working a double-shift*' [1]. It is, however, commonly accepted [1, 2, 10, 11, 15, 16, 19], that the structure of incentives for researchers has not favored the adoption of these costly practices. There is undeniably a cost of adhering to these principles that is '*not usually paid for by the funder nor supported by the home research institution (it may be seen as giving competing institutions an advantage) and it does not carry a premium*' [1].

The scientific community has identified this problematic structure of systemically created incentives and several initiatives have been initiated with the goal of mitigating this problem. Two of the most characteristic and impactive initiatives are the *Declaration on Research Assessment (DORA)* [20], and *The Hong Kong Principles for assessing researchers* [16], discussed below, in Sections 4.1 and 4.2 respectively. The goal of these initiatives is to motivate a systemic change of the incentive structure. Such a shift will reward those that follow best practices and will motivate others to do so as well. Moreover, it will provide researchers with the incentives to adhere to best practices without a negative impact on their careers.

4.1. DORA Declaration

The *Declaration on Research Assessment* (DORA) [20] was developed in 2012, during the Annual Meeting of the American Society for Cell Biology in San Francisco and ever since it has achieved a global impact across scientific disciplines. The goal of the declaration was to promote improved ways of evaluating researchers and scholarly research outputs. The declaration emphasizes the need to assess research on its own merits rather than based on the journal in which the research is published. The declaration has been signed by several institutions and individuals. The signatories of DORA support the adoption of the proposed practices in research assessment.

The Declaration urges funding agencies and research institutions to be explicit about the criteria they use in evaluating the scientific productivity of grant applicants and in taking hiring/tenure/promotion decisions respectively. More specifically, it suggests considering '*the value and impact of all research outputs (including datasets and software) in addition to research publications*' [20].
An example of the adoption of the DORA principles in practice is the policy of the Swiss National Science Foundation (SNSF). Since August 2020, the career funding schemes of SNSF have no longer considered the impact factors of scientific journals at any stage of the evaluation. SNSF takes a more holistic approach in evaluating the research output as a whole. '*This includes publications as well as other areas such as cooperation with stakeholder groups, science outreach, datasets, software, patents, conference papers and prizes*' [21]. This is an example of the concrete and tangible impact that initiatives like DORA can have in reformulating the incentive structure of academic research.

4.2. The Hong Kong Principles (HKPs)

The *Hong Kong Principles* (HKPs) for assessing researchers [16] were formalized in 2020, in the context of the 6th World Conference on Research Integrity. The focus of the initiative was on '*ensuring that researchers are explicitly recognized and rewarded for behaviors that strengthen research integrity*' [16]. More specifically, the HKPs emphasized the fact that '*research institutions should incentivize, reward, and assess individual researchers for behavior that fosters research integrity within their respective organization*' [16]. The abbreviated short version of the *Hong Kong Principles* can be summarized as follows:

(I) Assess responsible research practices

(II) Value complete reporting

(III) Reward the practice of open science

(IV) Acknowledge a broad range of research activities

(V) Recognize other essential tasks like peer review and mentoring

We will now discuss how these five principles overlap with the motivation of the current work. Firstly, Principle (I) suggests that authors who follow best practices that promote reproducibility and research integrity should receive the appropriate recognition in the research assessment process. It is underlined that following these practices comes at a cost, in terms of time and resources, and thus researchers abiding by best practices '*may disadvantage themselves compared to colleagues not participating in these practices*' [16]. Similarly, valuing complete reporting (Principle (II)) is linked to the fact that '*these activities deserve to be credited in the assessment of researchers because they are essential for replicability, to make it possible to verify what was done, and to enable the reuse of data*' [16].

Principe (III), '*Reward the practice of open science*', directly promotes '*open access, open methods, open data, open code*' [16] not only as facilitators of research reproducibility, but also as factors promoting equality to the research process. A strong argument in favor of open science is that '*a considerable amount of public funds is used for research, and its results can have profound social impact*' [16]. Moreover, in the spirit of openness, it is suggested that all participating research authors of a work should '*openly describe how each person has contributed to a research project*' [16], using for instance the CRediT taxonomy [13]. Moreover, the use of unique author identifiers, such as the Open Researcher and Contributor ID (ORCID), is proposed so that each researcher can be uniquely and unambiguously identified. Lastly, abiding by the FAIR principles [12] of data sharing is underlined as an appropriate practice.

Different types of research should be considered (IV): from creating new ideas and testing them to replicating key findings and synthesizing existing research. It is characteristically emphasized that '*replication studies or research synthesis efforts are often not regarded as innovative enough in* *researcher assessments, despite their critical importance for the credibility of research*' [16]. Lastly, activities like peer review should be recognized (V). Peer review is viewed as a cornerstone of research assessment and scientific progress. The quality of the review that researchers provide should also be a contributing factor in their assessment. Such contributions can be easily identified since it is common for journals to recognize reviewers' contributions, pointing towards these reviews using unique identifiers, in platforms such as *Publons* [22].

4.3. Funding Agencies

There is a growing interest over the subject of research reproducibility across scientific disciplines, among all types of stakeholders of scientific research. The growing awareness and realization of the stakes of encouraging transparency, openness and reproducibility in scientific research is driving policymakers to transform the current incentive structure. The policy changes are related to the preceding culture shift within the scientific community. As a result of the movement advocating for a more rigorous and fair research assessment, many funding agencies have taken significant steps in that direction.

A characteristic example is *Plan S* [23], which aims to enforce open-access publishing in the near future. *Plan S* was introduced in 2018, and it is supported by *cOAlition S*, an international consortium regrouping numerous national, international and charitable funders, as well as research organizations, under the support of the European Commission and the European Research Council (ERC). *Plan S* requires that '*from 2021, scientific publications that result from research funded by public grants must be published in compliant Open Access journals or platforms*' [23].

Steps in the same direction were also taken within the funding program Horizon 2020, of the European Union, which required research data of funded projects to be made available (with possible opt-outs), as well as demanding the submission of a '*Data Management Plan*' (DMP). Regarding its successor funding scheme, Horizon Europe (2021-2027), '*there are plans for compulsory DMPs and provisions in Model Grant Agreements (MGA) for open data availability*' [1]. In this scope, the report of the European Commission on the issue of the '*Reproducibility of results in the EU*' [1], makes concrete suggestions. More particularly, it suggests ensuring that '*issues related to research integrity are part of proposal evaluation*', and that '*reproducibility issues are part of DMPs*', making '*prior checks on existing results compulsory for research proposals*' and revising '*evaluation guidelines to reward robustness of methodologies*'. It is noteworthy that Horizon 2020 as well as Horizon Europe do not exclusively concern member-countries of the European Union, or associated countries, since they often allow the participation of researchers from third countries, a fact that underlines the global impact of the relevant decisions.

Similar initiatives have taken place in the USA as well, in which funders have directly linked reproducibility to evaluation. For instance, the National Institutes of Health (NIH) and the Agency for Healthcare Research and Quality (AHRQ) have '*put in place a policy, resources and training to support reproducibility, as part of their wider 'rigour' agenda*', including '*revised guidance concerning directly the evaluation of prior research in the instructions and review criteria for career development award applications and for Research Grant Applications*' [1].

There is, undoubtedly, a great ongoing systemic shift of national and international organizations towards the support of the adoption of best practices, which favor open and reproducible science. These practices have started becoming formal funding requirements as well as elements of the evaluation for research funding. Within this new emerging structure of incentives and legal requirements, it becomes evident that the sooner individual researchers and research communities adapt to these requirements, the faster their research will move forward.

4.4. Benefits for the Authors

It is understandable that requiring researchers to dedicate additional effort to satisfy the requirements of openness, transparency and reproducibility, might be viewed with hesitation, concern or unwillingness. Building on the argumentation presented in previous sections, we now enumerate ten concrete benefits (proposed by Gundersen et al [11]) that may motivate the authors to put in the additional effort that is, undeniably, required.

1) Contribute to the promotion of more rigorous science
2) Receive credit for all research output (datasets, code)
3) Increase visibility and citability of your research
4) Better funding potential under new requirements
5) Offer variety to your CV
6) Improve management of your research assets
7) Facilitate the reproduction of your work
8) Timely adaptation to new publishing requirements
9) Attract transformative students and colleagues
10) Demonstrate leadership and forward thinking

In addition to these benefits for the individuals, it is important to also reflect on the positive repercussion on the research community of a certain field, such as indoor positioning:

(i) Field moves forward faster (direct reuse of existing methods, baselines).
(ii) Ability to make consistent and repeatable comparisons and to select the best performing methods.
(iii) Gaining time by not reinventing the wheel, but focusing on more meaningful research directions.
(iv) Have the infrastructure to be competitive as a discipline in the forthcoming research funding context.
(v) Transparent overview of the State of the Art, facilitating the market adoption of the most adequate methods.
(vi) The field can become renowned for being at the forefront of the effort for research reproducibility.

5. Suggestions for the Indoor Positioning Community

The presented initiatives of the scientific community, of the related funding agencies and of other involved stakeholders, suggest that this emerging culture of transparency and openness of the scientific process is moving on from being a noteworthy and rare practice, to becoming a *sine qua non*. Apart from the cultural dimension, concrete policy measures have already started being implemented in the research funding process, across all research disciplines. For some disciplines it might be inherently more difficult to support reproducibility, compared to others. The field of indoor positioning research enjoys a head start compared to other disciplines, due to the nature of the discipline and to the fact that the vast majority of the research outcomes can be presented and shared in the form of computer code. Moreover, certain established activities of the community, such as the IPIN competition [24], constitute an excellent example of transparent, rigorous and consistent comparison of research outcomes. In this section, we propose a non-exhaustive list of concrete steps that the indoor positioning research community could take to accelerate its pace towards becoming a truly reproducible discipline. These steps could be encouraged by the IPIN conference, as well as by other related conferences, journals and special sessions that lie within the thematic area of indoor positioning.

5.1. Reproducibility Checklist

A first idea that could be easily implemented and that could have a direct and significant repercussion, would be the creation of a checklist of reproducibility-enhancing points for future paper submissions. The suggested points, which are in line with the ongoing similar actions, are:

(1) **Open code**: Do the authors provide the code implementation related to the paper's content? Are all relevant pieces of information, that would facilitate the reproduction of the tests provided (as for instance, versions of packages/libraries used, dependencies declared, etc.)? If the code implementation

is not shared, is there a justification provided (Reasons for opting out could be: IPR issues, insufficient resources, third-party owning part of the code, etc.)?

(2) **Open FAIR data**: If the authors use data that they collected, do they share these data? Does the data sharing comply with the FAIR principles? Is the data collection method sufficiently described (method of collection, system for defining the spatiotemporal ground truth and its estimated level of accuracy, etc.)? Is the train/validation/test set repartition of the data available? If the dataset used is not shared, is a justification provided (Reasons for opting out could be: IPR issues, insufficient resources, third-party owning the data, etc.)? If the authors use public data, is the source clearly referenced?

(3) **Deployment description**: If a deployment is used in the paper, is it sufficiently described (Ground truth collection method, type(s) of Access Point (AP) technology used, type of mobile devices, AP density, type/map of environment, size of the area of AP deployed, data collection area size, etc.)?

(4) **Evaluation protocol description**: Is the evaluation protocol clearly described? Is there a baseline method used? Is the proposed method compared against the baseline in a fair and consistent way (using, for instance, the same dataset)?

(5) **Declared roles of contributors**: Are the roles of all contributors appearing in the author list clearly defined (A simple adoption of the *CRediT* [13] author statement can address this point)?

The above proposed items, or a subset of them, could be adopted, either as optional or as mandatory points that a paper submission should fulfil. The level of compliance with these points could also become a factor of the evaluation of papers. If such a checklist were to be adopted, its requirements would need to be publicly announced in the call for papers. Moreover, a mechanism such as the *Open Science Badges* [25], could be a first step of positive encouragement of these best practices.

5.2. Other Actions

In addition to the above discussed checklist, there exist other simple steps that the community could take, which could encourage and motivate actions towards more reproducible research. Indicatively, we propose:

(i) The adjustment of a more flexible structure of **Peer Review** questionnaires,
(ii) The encouragement of **Replication Studies** in a formal context,
(iii) The adoption of the **Registered Reports** format,
(iv) The repeating evaluation, through time, of the level of reproducibility of the field by **Scoping Studies** and surveys,
(v) **Recognizing and rewarding** outstanding efforts facilitating reproducibility.

These actions are becoming common practices in many disciplines, and they could be particularly useful in the field of indoor positioning research. More specifically, in relation to the first point, the questions that a reviewer has to answer when assessing a paper submission in the context of the **Peer Review** process (i), are often written in a way that predetermines that submissions with positive results will be preferably evaluated. This may result in unfairly treating other types of studies (studies with negative results, replication studies, surveys) which should be encouraged as well, and should be evaluated with criteria that are relevant to their nature. Thus, the questions that the reviewers are invited to answer could be enriched or could be made adaptable to the type of the paper, without favoring a-priori novelty over replication. Depending on whether the paper concerns a novel idea (positive results), a replication study (negative or confirmatory results), or a survey (literature review / State of the Art), the peer review procedure could be adapted accordingly.

The idea of **Replication Studies** (ii) can be a very interesting direction for the indoor positioning research community. A common experience of experts of this field is the fact that the fine-tuning of an IPS at each distinct deployment is a very important task that is determinant for the system's performance. Inviting the community to replicate published works and to evaluate the consistency of results, either on the same or on different datasets, is a precious exercise to encourage. The invitation for replication studies could take place in the context of special sessions at conferences like IPIN, on special issues in relevant journals or simply as a special submission type. Such submissions should respect some minimum requirements like open code and open data, both with a persistent identifier. The precious material that the community possesses from the annual IPIN competitions, could be used as excellent starting points for such studies. The submitted solutions of a certain year could be tested with data from upcoming years, upon the necessary adjustments that might be required. Such studies would evaluate the consistency of the performance of competing systems in different settings and would showcase the generalization potential of these systems.

The concept of **Registered Reports** (iii) could function as an indispensable steppingstone for the indoor positioning community in its effort to establish standards based on community agreement and, eventually, on consensus. Submitting for peer review the protocol and the methodology that is intended to be used in large scale experiments or in data collection projects, can offer a precious, timely, and targeted feedback by fellow peers, which can improve the overall design of the planned work. Moreover, should a registered report be accepted, its authors are reassured that the result of their laborious planned work will be accepted for publication. The potential for peer intervention and discussion on the protocol design phase facilitates the execution of protocols with a wider acceptance.

Registered Reports would be particularly useful for the cases of data collection projects, which aim to collect and publicize datasets. In such cases, a community approval of the data collection protocol before the actual data collection, gives the community the opportunity to control the quality of such processes. Since the public datasets are widely used by numerous research works of the field, the quality of their collection protocol may have a tremendous repercussion in the conclusions drawn by all their citing works. The timely feedback of peer review facilitated by the Registered Reports can greatly affect the quality of data collection works, and subsequently, of all future works reusing their materials.

The scoping report of the European Commission on the issue of the *Reproducibility of results in the EU* [1], indicates that '*there is a growing awareness of the problem in many disciplines, testified by scoping studies and seminal surveys*'. Indeed, interesting **Scoping Studies** (iv) in various disciplines [11, 26, 27, 28, 29] depict a realistic picture of the status of reproducibility in their respective field, following rigorous examinations of the relevant literature and providing fact-based conclusions. An objective understanding of the status of a field can assist the decision-making towards positive change. Therefore, a relevant survey on the current level of reproducibility of indoor positioning research would be an excellent steppingstone, and a future baseline to evaluate the progress of the indoor positioning community in this aspect. Lastly, such works are excellent examples of scenarios where the logic of Registered Reports would be appropriate, as the protocol of assessing the level of reproducibility of previous works of a field would be firstly peer reviewed and approved.

Recognizing and rewarding (v) researchers for their outstanding efforts to facilitate reproducibility is an efficient and inexpensive way in which the community could promote the culture of reproducible research [30]. One such example are the awards; along the various existing types of awards (best paper, best student paper, best presentation, best poster, etc.) '*awards, such as for outstanding effort to make complex results more reproducible or outstanding effort to reproduce results*' [30] could be added. Moreover, relevant conferences or journals could provide a definition of what constitutes '*an article with reproducible results and recognize these articles (e.g., with badges or other incentives)*' [30]. Lastly, the comparison of reproducible results across articles could take place in the form of a competition [30].

6. Conclusions and Future Work

In this work, we described the advancements in the world of academic science towards more reproducible research. The particular viewpoint was that of the Indoor Positioning research community. We proposed a series of steps that the community could undertake to assist *Research Reproducibility*. Moreover, the benefits of such actions within the transforming landscape of incentives in the academic world were extensively presented. In this context, we consider that most of the proposals are relatively inexpensive to implement.

Overall, there is a debate whether there is a '*replication crisis*' in science. Nevertheless, Reproducibility should not be framed as a crisis, but rather as an '*ideal*' [1]. This approach may summarize the mentality with which the numerous presented initiatives have motivated the ongoing culture shift. Indoor positioning research is at a crossroads. The community possesses the capital of several public datasets, which are being published at an increasing pace. Established activities of rigorous comparisons, such as the IPIN competitions, are part of the richness of the field. Nevertheless, the community has yet a lot of important steps to undertake. Directions such as increased openness of data and code, complete and unambiguous reporting, and standardized evaluation processes require urgent action. These directions, if followed, would provide the push forward that the field deserves.

As a part of their immediate future plans, the authors of this work intend to design a protocol for studying a volume of works of the field and creating a survey that will provide an overview of the *status of reproducibility of the field*. It would be ideal if such a work were conducted with the wider possible participation of experts in this field, who are warmly invited to co-create it.

7. Acknowledgements

This work was funded by the Swiss National Science Foundation, under the *Spark* Funding scheme, in the context of the project *Eratosthenes* (No. 195964).

The first author would like to acknowledge the *Swiss Reproducibility Network* (SwissRN), since his participation to it has inspired the current work and has facilitated the access to multiple interesting relevant resources.

8. References

[1] W. Lusoli, Reproducibility of Scientific Results in the EU: Scoping Report. Publications Office of the European Union, 2020. [Online].

[2] National Academies of Sciences, Engineering, and Medicine, Reproducibility and Replicability in Science. Washington, DC: The National Academies Press, 2019. [Online]. Available: https://www.nap.edu/catalog/25303/reproducibility-and-replicability-in-science

[3] H. E. Plesser, Reproducibility vs. Replicability: A Brief History of a Confused Terminology, Frontiers in neuroinformatics, vol. 11, pp.76–76, Jan. 2018, publisher: Frontiers Media S.A. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29403370

[4] L. A. Barba, Terminologies for reproducible research, 2018

[5] B. T. Essawy, J. L. Goodall, D. Voce, M. M. Morsy, J. M. Sadler, Y. D. Choi, D. G. Tarboton, and T. Malik, A taxonomy for reproducible and replicable research in environmental modelling, Environmental Modelling & Software, vol. 134, p. 104753, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364815219311612

[6] S. Adler, S. Schmitt, K. Wolter, and M. Kyas, A survey of experimental evaluation in indoor localization research, in 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2015.

[7] F. Potortì, S. Park, A. R. Jiménez Ruiz, P. Barsocchi, M. Girolami, A. Crivello, S. Y. Lee, J. H. Lim, J. Torres-Sospedra, F. Seco, R. Montoliu, G. M. Mendoza-Silva, M. D. C. Pérez Rubio, C. Losada-Gutiérrez, F. Espinosa, and J. Macias-Guarasa, Comparing the performance of indoor localization systems through the evaal framework, Sensors, vol. 17, no. 10, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/10/2327

[8] C. M. de la Osa, G. G. Anagnostopoulos, M. Togneri, M. Deriaz, and D. Konstantas, Positioning evaluation and ground truth definition for real life use cases, in 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2016.

[9] G. M. Mendoza-Silva, J. Torres-Sospedra, F. Potortì, A. Moreira, S. Knauth, R. Berkvens, and J. Huerta, Beyond euclidean distance for error measurement in pedestrian indoor location, IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–11, 2021.

[10] M. R. Munafò, B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. Percie du Sert, U. Simonsohn, E.-J. Wagenmakers, J. J. Ware, and J. P. A. Ioannidis, A manifesto for reproducible science, Nature Human Behaviour, vol. 1, no. 1, p. 0021, Jan. 2017. [Online]. Available: https://doi.org/10.1038/s41562-016-0021

[11] O. E. Gundersen, Y. Gil, and D. W. Aha, On reproducible AI: Towards reproducible research, open science, and digital scholarship in AI publications, AI Magazine, vol. 39, no. 3, pp. 56–68, Sep. 2018. [Online]. Available: https://ojs.aaai.org/index.php/aimagazine/article/view/2816

[12] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The FAIR guiding principles for scientific data management and stewardship," Sci Data, vol. 3, p. 160018, Mar. 2016.

[13] "CRediT - contributor roles taxonomy," https://casrai.org/credit/, Sep. 2019, accessed: 2021-5-5.

[14] Center for Open Science, "Registered reports," https://www.cos.io/initiatives/registered-reports, accessed: 2021-5-7.

[15] R. A. Lundwall, Changing institutional incentives to foster sound scientific practices: One department, Infant Behavior and Development, vol. 55, pp. 69–76, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0163638318300900

[16] D. Moher, L. Bouter, S. Kleinert, P. Glasziou, M. H. Sham, V. Barbour, A.-M. Coriat, N. Foeger, and U. Dirnagl, The hong kong principles for assessing researchers: Fostering research integrity, PLOS Biology, vol. 18, no. 7, pp. 1–14, 07 2020. [Online]. Available: https://doi.org/10.1371/journal.pbio.3000737

[17] A. Cockburn, P. Dragicevic, L. Besançon, and C. Gutwin, Threats of a replication crisis in empirical computer science, Commun. ACM, vol. 63, no. 8, p. 70–79, Jul. 2020. [Online]. Available: https://doi.org/10.1145/3360311

[18] B. A. Nosek, J. R. Spies, and M. Motyl, Scientific utopia: II. restructuring incentives and practices to promote truth over publishability, Perspectives on Psychological Science, vol. 7, no. 6, pp. 615–631, 2012, pMID: 26168121. [Online]. Available: https://doi.org/10.1177/1745691612459058

[19] N. Mejlgaard, L. Bouter, G. Gaskell, P. Kavouras, N. Allum, A.-K. Bendtsen, C. Charitidis, N. Claesen, K. Dierickx, A. Domaradzka, A. Elizondo, N. Foeger, M. Hiney, W. Kaltenbrunner, K. Labib, A. Marušić, M. Sørensen, T. Ravn, R. Scepanovic, and G. Veltri, Research integrity - nine ways to move from talk to walk, Nature, vol. 586, 10 2020.

[20] "Declaration on research assessment (dora)," https://sfdora.org, Jan. 2018, accessed: 2021-4-30.

[21] "The snsf has signed the dora declaration," http://www.snf.ch/en/theSNSF/research-policies/dora-declaration/Pages/default.aspx, accessed: 2021-4-30.

[22] "Publons.com - the home of expert peer review." https://publons.com/about/home, accessed: 2021-5-4.

[23] "Plan s - an initiative for open access publishing," https://www.coalition-s.org, accessed: 2021-5-4.

[24] J. Torres-Sospedra, A. R. Jiménez, A. Moreira, T. Lungenstrass, W.-C. Lu, S. Knauth, G. M. Mendoza-Silva, F. Seco, A. Pérez-Navarro, M. J. Nicolau, A. Costa, F. Meneses, J. Farina, J. P. Morales, W.-C. Lu, H.-T. Cheng, S.-S. Yang, S.-H. Fang, Y.-R. Chien, and Y. Tsao, Off-Line Evaluation of Mobile-Centric indoor positioning systems: The experiences from the 2017 IPIN competition, Sensors, vol. 18, no. 2, Feb. 2018.

[25] Center for Open Science, "Open science badges", https://www.cos.io/initiatives/badges, accessed: 2021-5-5.

[26] J. H. Stagge, D. E. Rosenberg, A. M. Abdallah, H. Akbar, N. A. Attallah, and R. James, Assessing data availability and research reproducibility in hydrology and water resources, Scientific Data, vol. 6, no. 1, p. 190030, Feb. 2019. [Online]. Available: https://doi.org/10.1038/sdata.2019.30

[27] O. E. Gundersen and S. Kjensmo, State of the art: Reproducibility in artificial intelligence, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1, Apr. 2018. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/11503

[28] N. Bonneel, D. Coeurjolly, J. Digne, and N. Mellado, Code replicability in computer graphics, ACM Trans. Graph., vol. 39, no. 4, Jul. 2020. [Online]. Available: https://doi.org/10.1145/3386569.3392413

[29] O. B. Amaral, K. Neves, A. P. Wasilewska-Sampaio, and C. F. Carneiro, The Brazilian Reproducibility Initiative, eLife, vol. 8, p. e41602, Feb. 2019, publisher: eLife Sciences Publications, Ltd. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30720433

[30] D. E. Rosenberg, Y. Filion, R. Teasley, S. Sandoval-Solis, J. S. Hecht, J. E. van Zyl, G. F. McMahon, J. S. Horsburgh, J. R. Kasprzyk, and D. G. Tarboton, The next frontier: Making research more reproducible, Journal of Water Resources Planning and Management, vol. 146, no. 6, p. 01820002, 2020. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29WR.1943-5452.0001215