

“We provide our resources in a dedicated repository” Surveying the Transparency of HICSS publications

Irdirin Pekaric
University of Liechtenstein
Vaduz, Liechtenstein
irdin.pekaric@uni.li

Giovanni Apruzzese
University of Liechtenstein
Vaduz, Liechtenstein
giovanni.apruzzese@uni.li

Abstract

Every day, new discoveries are made by researchers from all across the globe and fields. HICSS is a flagship venue to present and discuss such scientific advances. Yet, the activities carried out for any given research can hardly be fully contained in a single document of a few pages—the “paper.” Indeed, any given study entails data, artifacts, or other material that is crucial to truly appreciate the contributions claimed in the corresponding paper. External repositories (e.g., GitHub) are a convenient tool to store all such resources so that future work can freely observe and build upon them—thereby improving transparency and promoting reproducibility of research as a whole.

In this work, we scrutinize the extent to which papers recently accepted to HICSS leverage such repositories to provide supplementary material. To this end, we collect all the 5579 papers included in HICSS proceedings from 2017–2024. Then, we identify those entailing either human subject research (850) or technical implementations (737), or both (147). Finally, we review their text, examining how many include a link to an external repository—and, inspect its contents. Overall, out of 2028 papers, only 3% have a functional and publicly available repository that is usable by downstream research. We release all our tools.

Keywords: Research, Transparency, Replicability, Reproducibility, Repository, Resources, Open Science

1. Introduction

The Hawaii International Conference on System Sciences (HICSS) is one of the largest outlets for scientific advances. Since 1967, HICSS represents a hub for researchers from a plethora of domains gravitating around information technology (IT) and

social sciences. For instance, recently accepted works encompass digital innovation (Nijse & Litchfield, 2023), law and regulation (Koh et al., 2024), software development (Heine et al., 2024), or business intelligence (Ul-Ain et al., 2019). The “numbers” of HICSS are constantly increasing (e.g., 550 papers in 2012, 635 in 2017, and 764 in 2024), indicating that, every year, more and more cutting-edge discoveries are made and are significant enough to be discussed.

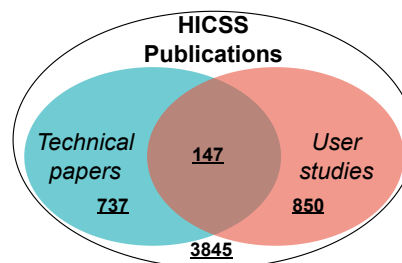


Fig. 1. Distribution of the papers analysed in our research – From 5579 papers published at HICSS since 2017, we isolate 884 technical papers and 997 user-studies papers (147 papers are shared), and examine their repositories (if any).

The growth seen by HICSS reflects the generic trend of research (To & Yu, 2023). Unfortunately, despite these successes, the rising number of scientific publications may conceal various problems—among which one, in particular, stands out: the **reproducibility crisis** (Baker, 2016). Simply put, even though novel findings are published every new day, researchers are having troubles reproducing the results of prior work. This phenomenon is claimed to affect various research fields—most notably, artificial intelligence (AI), but also biology or business (Aguinis et al., 2020; Kapoor & Narayanan, 2023; Miyakawa, 2020). Such “endemicity” is problematic, since it casts doubts on the conclusions of researchers, undermining their credibility.

Among the root causes of the reproducibility crisis,

there are two culprits: the poor **replicability** (Plesser, 2018) and **transparency** (Hardwicke et al., 2020) of research papers. Indeed, sometimes papers may not include the information required to even attempt reproduction of the experiments described therein. For instance, some low-level details may not be provided, the implementation of an algorithm may not be disclosed, or the data to test a hypothesis may not be released. Altogether, these “lacks” impair reproducibility, since downstream researchers are likely to introduce some deviations in their setup when reproducing a prior work—leading to different results.

Importantly, these “omissions” may be due to valid reasons, such as strict page limit, data confidentiality, or proprietary/patented software (Apruzzese et al., 2023; Pekaric et al., 2021). However, while the latter reasons are unavoidable, there exist ways to address the former: external **repositories**. For instance, GitHub is a platform wherein any sort of supplementary research material (e.g., documentation, data, source code) can be uploaded (Milliken et al., 2021), thereby representing a solution to the limited space available on written manuscripts. Still, despite the existence of such tools (which could mitigate the reproducibility crisis and improve transparency of research), their adoption is not widespread yet: to provide a recent example, Olszewski et al. (2023) showed that barely half of the publications in top-tier cyber security venues release their artifacts.

We observe that no prior work attempted to analyse the adoption of such repositories for HICSS. Hence, **we ask ourselves**: “*To what extent do HICSS papers release their resources into publicly accessible repositories?*” Such a research question (RQ) serves to understand the overall transparency of HICSS’ papers, thereby allowing to: (i) gauge their replicability and, therefore, the reproducibility of their results; and (ii) establish a foundation for future work, facilitating methods re-use. Moreover, we dissect our main RQ into two sub-RQ by decoupling the term “HICSS papers” into two distinct terms: “user studies” (RQ1) and “technical papers” (RQ2). Our choice is motivated by the better utility that public repositories have for these classes of papers.

CONTRIBUTIONS. We want to emphasize the role of external repositories linked within HICSS publications. After defining the problem space in Section §2:

- First, through a custom-developed tool, we collect all papers accepted to HICSS in 2017–2024 (§3.1);
- Then, we rigorously review all these works and selectively extract “user studies,” entailing human-subject research; and “technical papers,” entailing implementation of code or tools (see Fig. 1);
- Finally, we scrutinize the extent to which such works publicly share their artifacts onto (currently active) external repositories, and inspect their contents (§3.3).

We present our findings (§4) and draw implications for future work (§5). For the sake of open science, we release our resources at: <https://github.com/hihey54/hicss58/>

2. Background and Preliminaries

We position our work within extant literature, and then define our scope and explain its relevance.

2.1. Related Work

Abundant prior work has investigated the theme of transparency or reproducibility of research papers.

Oftentimes, such analyses consider papers published in high-quality venues to provide more compelling findings. For instance, Apruzzese et al. (2023) analysed 88 papers published in top-4 security conferences focusing on “adversarial machine learning”, and found that only half release their source code. Similarly, Olszewski et al. (2023) specifically focus on the “artifacts” related to top-tier security venues, investigating whether they run as described in the respective publication: out of 298 artifacts, only 20% produce the same results.

By adopting a less-technical viewpoint, Daneshvar Kakhki et al. (2021) consider 82 user surveys within nine top journals in information systems, and found that not one provides enough details for replication. Another perspective is given by Miyakawa (2020), who found that, out of 180 papers submitted to a well-known journal in biology, 41 were explicitly asked to provide raw data before the beginning of the reviewing process: among these, only 1 provided such data. The recent study by Fišar et al. (2024) considered 500 articles in a respected management journal, showing that data is one of the most limiting factors in reproducibility. Intriguingly, Serra-Garcia and Gneezy (2021) even claim that non-replicable publications are cited more than replicable ones.

We observe that our work also pertains to that branch of research that focuses on how the world embraces innovations (Davis et al., 1989; Rogers, 2003). For instance, many studies, such as Taherdoost (2018), have investigated how researchers react to progress in the context of information technology. Closely related to our study is the work by Escamilla et al. (2022), revealing an increased usage of software repositories among academics—according to papers appearing on arXiv (which are not necessarily peer-reviewed) and PubMedCentral (focusing on medicine, which is outside our scope). Nonetheless, in carrying out our analysis, we will borrow the methodological approach (in terms of scripting) adopted in Escamilla et al. (2022).

To sum up, prior work showed that transparency and reproducibility issues may affect a range of domains. Such a phenomenon justifies our choice to focus on HICSS, which is host to cutting-edge and peer-reviewed publications from a constellation of research fields: hence our results can be inspirational for future work.

2.2. Definitions and Focus of the paper

HICSS papers encompass various types of contributions. We focus on two classes of works:

- *Technical papers.* These articles rely on a strong technical component to test their hypotheses. Examples include: papers that develop an original system/tool (Giudici et al., 2024); papers that use existing tools to carry out some analyses (Majchrzak et al., 2022), potentially with some additional scripts (Scherer et al., 2022). Obviously, we do not consider as “technical” a paper which relies on trivial software (e.g., a desktop, or a text editor).
- *User studies.* These articles entail human-subject research. Examples include: user surveys (Sun et al., 2021); behavioral studies (Majchrzak et al., 2022; Yan & Gurkan, 2023); or interviews (Nägele et al., 2024).

We thoroughly explain the reason for choosing these classes of works in the next subsection. Notably, some papers may fall into both categories: this is typically the case for “design science research”, wherein interviews or surveys are carried out before developing a certain tool, which is then developed and tested by some users to draw conclusions (Nguyen et al., 2020).

Importantly, some papers may not fall in either category—this is common for literature reviews (Ul-Ain et al., 2019), or conceptual frameworks (Asprion et al., 2019). The choice to exclude these works is that they have a reduced necessity to upload their resources to repositories (since they typically have no evaluation), and they are hence likely to be self-contained. Therefore, it would be unfair to consider similar works in our analysis, as it would unfairly skew the results—since we seek to analyse HICSS as a whole.

2.3. Motivation of our Design Choices

The common element of the papers discussed in §2.1 was that they all focused on a single domain. Here, however, we must account for the huge variety of papers (and corresponding methods) that are presented at HICSS. Therefore, to carry out a humanly feasible analysis, we will assess whether any given paper (among the classes we consider) provides the link to an active external repository. Let us explain why this is important.

For technical papers, external repositories are crucial. According to our definition, it is impossible

to provide all elements (e.g., code, data, parameter configurations) necessary to accurately reproduce the results in a single manuscript—especially given that HICSS papers are typically 10-pages long without any appendices. Hence, the lack of an external repository may impair the reproducibility of any technical paper.

For user studies, external repositories are also fundamental. For instance, it may be useful if, for interviews, one could release the answers (potentially anonymised). For surveys, sharing the questionnaire could be ideal to inspect how the various questions are presented to the user: evidence suggests that even the ordering of questions may change the responses of surveys (Tourangeau et al., 2004). Given that it is unlikely for all such elements to be provided in the paper, existence of repositories is paramount also for this class of works—from a transparency perspective.

The reasons above motivate our choice of focusing on our considered RQs. Moreover, we also investigate *what is included* in the repositories linked in such prior work. Our paper seeks to answer all such questions.

3. Research Method

We describe the methodology followed for our analysis.

3.1. Data Collection

The starting point of our contributions is the collection of HICSS publications spanning 2017–2024; we consider this timespan because HICSS proceedings have been hosted (for free) on [ScholarSpace](#) since 2017. We collect such a dataset by developing a custom *scraper*. We provide the code of our scraper in our repository.

Our scraper iteratively downloads HICSS papers from the proceedings web page, organized as follows. From the starting page, it is possible to select a specific edition of HICSS. Then, we are brought to the main tracks of that edition. From here, it is possible to choose the various mini-tracks of the considered track. Finally, upon selecting a mini-track, it is possible to visualize and download the papers included in that mini-track.

We implement our scraper by following the procedure above. Overall, we downloaded a total of 5579 publications (whose cumulative size is ≈5.42GB). Notably, this number matches the one provided by summing all the numbers of accepted papers of HICSS as mentioned in their preface (i.e., 5579). This confirms the quality of our scraper, and the soundness of our methods. The resulting dataset has the same organization described above (years-tracks-minitracks-papers) to facilitate navigation.

<p>Ethics: We cannot publicly release our dataset. However, by using our scraper, one can collect the same dataset: the HICSS’ proceedings are publicly available, free of charge.</p>

3.2. Preliminary (automatic) Paper Analysis

The second step entails analysing the 5579 collected publications. However, given their large number, we carry out a preliminary analysis in an automated way by developing a *custom script* (provided in our repository), as also done by Escamilla et al. (2022).

At its core, our script iteratively inspects the content of each paper (a .pdf file) in our dataset. For every paper, the script is designed to answer three questions: “does the paper entail a user study?”, “does the paper involve some technical implementation?”, “does the paper provide a link to a dedicated repository?”. To answer each of these questions, the script looks for certain keywords included in a set of **three lists** which we autonomously devised. Specifically:

- *user study*: variations of (user study, questionnaire, online survey, user survey, interview);
- *technical paper*: variation of (empirical, experiment, evaluation, source code, artifact, implementation, tool);
- *repositories*: variations of (github.com, gitlab.com, zenodo.org, figshare.com, anonymous.4open.science, drive.google.com, onedrive.live.com, 1drv.ms).

The exact lists, and corresponding patterns, are observable in the code of our script. We derived these lists after having qualitatively inspected some papers, but also through our own domain expertise.

Hence, we let the script run and process all papers included in our dataset. To **identify potential candidates** for the “user study” or “technical paper” categories, the respective lists of keywords are checked *only in the abstract*: if any match is found, the paper is considered as a candidate for the corresponding category. We only check the abstract because our results would be too noisy and unreliable if we considered the whole paper. Importantly: the search does not stop at the first occurrence. Therefore, the script is also likely to assign the same paper to both categories. Simultaneously, the script also checks the *entire content of the paper* for patterns mimicking the corresponding list that could point to external resource repositories.

At the end of this preliminary procedure, we obtain a set of 1949 and 551 papers: the former being potential candidates for technical papers (of which 133 may have a dedicated repository), the latter for user studies (of which 11 may have a dedicated repository); 173 papers are placed in both categories. 3425 papers are not included in any category, and will not be considered in the following analysis.

3.3. Manual Review and Validation

It is unrealistic to expect that a simple and fully-automatic keyword check guarantees accurate

extraction of the information we need to answer our RQs. Hence, as the last step of our procedure, we manually review the papers in each group of candidates. The intention is threefold: (i) ensure that the papers are included in the correct category; (ii) verify the paper has a link to its specific repository, and (iii) check the content of the repository—if available.

Such an analysis is **done in pairs**, in a similar fashion as done by Apruzzese et al. (2023): two authors work independently to inspect the papers and assign the category to each of them. The authors frequently discussed their findings, and if there were any uncertainties about their judgements, they consulted each other to reach a consensus.¹ These analyses are qualitative in nature, but follow a rigorous approach encompassing the entire text of the paper. For instance, for “user study”, we used manual keyword search (with similar terms as in the corresponding list), but also visual inspection (e.g., tables or figures); for “technical papers”, we examine if non-trivial software² has been used during the study—either by using existing tools with minimal changes, or modifying previous scripts, or developing new systems; whereas for “repositories”, we check if the links point to a repository that is truly associated to the activities carried out in the paper (e.g., links pointing to repositories created beforehand, or from different authors, are not counted).

To provide more insightful results, we not only scrutinize if a paper is a user study or a technical paper, but **we delve deeper**. For technical papers, we are particularly interested in the nature of the provided repositories, asking ourselves “do these repositories include data, code, or an entire system?”. For user studies, we first check if the paper is an observational study, or a user survey, or interviews; and then explore their repositories, checking whether they include supplementary material that is related to the nature of the user study (e.g., questionnaires). Hence, we manually check the content of all repositories associated with any given paper included in our analysis.

We show a summary representation of the results of our manual analysis in Fig. 1. Before addressing our RQs, we find it instructive to determine the **quality of our custom script**. We do so by measuring how many papers, after our manual inspection, had been correctly classified by our script. For user studies, our script identified 551 papers, but our manual check found 997

¹During the reviewing phase of this paper (hence *after* we had carried out our analysis) we have carried out an experiment to measure the **inter-coder reliability score**. Each author randomly chose 50 papers that they independently reviewed, and asked the other author to assign the corresponding class. We observed an agreeability of 93%, denoting that the authors likely reach the same conclusion.

²We consider a “trivial software” as those programs that are a prerequisite to carry out research activities, e.g., MS Word for writing, or Excel for analytics or visualizations, or PowerPoint for slides.

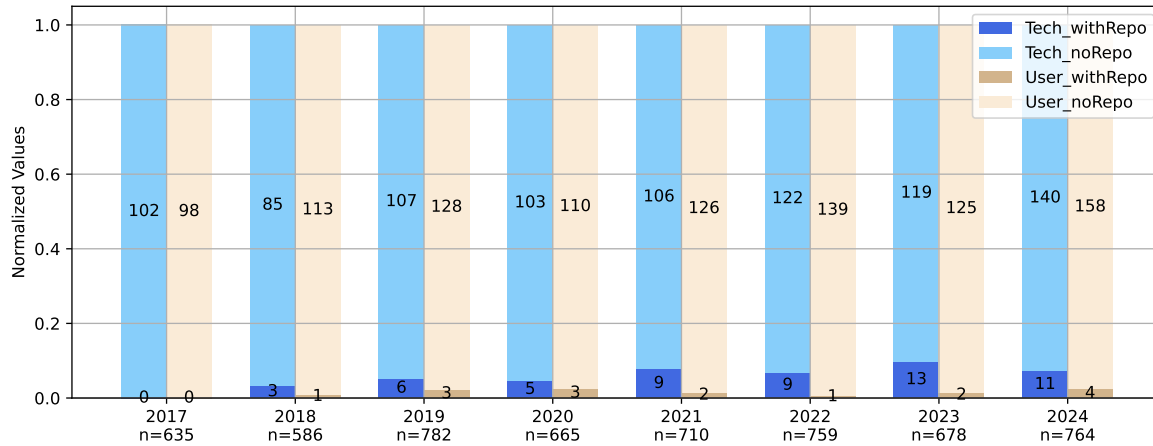


Fig. 2. Temporal distribution of our analysed papers – The numbers inside the bars are the absolute numbers, while the bars represent the proportion of the papers (user studies and technical) that have (or not) functional repositories. Ticks on the x-axis report the year and the number of papers accepted to HICSS for that year.

(some of these were identified in the set of technical papers); for technical papers, our script identified 1949 papers, but we found 884; for matching repositories, our script identified 144, but we found 71. Hence, we endorse downstream researchers to *exercise caution* when using our script to perform similar analyses, as the number of misclassifications is substantial.

4. Results

We now present the results of our research after our manual validation. To this end, we first provide in Fig. 2 the temporal distribution over the considered timespan (2017–2024) of our findings. This plot shows how many papers for each class have a functional link³ to a corresponding external repository (papers falling in both classes are counted in each class). Specifically, the x-axis shows the years, and for every year the number of total papers accepted at HICSS; the y-axis reports the relative distribution of the papers with/without an external repository; numbers on the bars denote the absolute number of papers. We can already see that the overall percentage of papers with repositories is abysmally low ($\approx 3\%$). However, let us further analyse our findings by focusing on our RQs.

4.1. User Studies (RQ1)

We first focus on papers entailing user studies.

First, out of 997 papers entailing user studies, only 16 (1.6%) provide supplementary resources in an external repository. According to Fig. 2, it is evident that this number ranges from 0 to 4 for each of the eight years. In the year 2017, there were no user

³By “functional link” we mean that it leads to a repository that is active and has some content, and hence serves a “function”.

studies that linked external repositories. However, going forward it is visible that this number started to increase. As a result, in the years 2019 and 2020, 3 studies in each year provided their supplementary materials in linked repositories. In this regard, the last year (2024) had the highest number (4), which can be considered encouraging and could indicate that the growing trend will continue. Moreover, it is also noticeable that the number of user studies tends to increase every year (also w.r.t. HICSS overall papers): 98 papers (out of 635) included user studies in 2017, 129 (out of 665) in 2021, 140 (out of 759) in 2022 and 162 (out of 764) in 2024.

Second, delving deeper, we found that, out of 997 papers having user studies, 472 (47%) entail surveys, whereas 595 (60%) interviews and only 42 (4%) are behavioral studies.⁴ In regards to user surveys, 11 studies (2%) provide the URL to the external repository. The numbers do not differ extensively for interviews (7, i.e., 1%) and behavioral studies (1, i.e., 2%). This is presented in Table 1.

Third, focusing on the contents of the provided repositories, we found that accessing repositories can also be a problem. Out of 16 repositories, we were able to access 13 (81%) repositories. There are two reasons for this: either the *URL is broken*, or the *repository is private* (we do know if these issues are only occurring at the time of our analysis in June 2024, or have always been present). This can make it difficult to evaluate any shared contents as well as to reproduce the conducted studies. Furthermore, we checked the contents of the 13 accessible repositories. The shared materials included various artifacts such as surveys, interview descriptions, codebooks, questions, data, results, code, and tools. It is

⁴We note that the sum of these percentages exceeds 100% because multiple papers contained “combinations” of surveys and interviews (sometimes even behavioral studies).

important to note that the last two mentions relate to user studies that could be also considered technical papers.

Finally, we found that user studies represent 19% of the overall number of HICSS publications, emphasizing the role of user-centered research in the field of information systems and technology. This is further underscored by the fact that it is growing every year, indicating a sustained commitment to studies involving human subject research. We also found that a small percentage (2.6%) of user studies can be classified also as “technical papers”, a classification that is particularly popular for some more technical tracks at HICSS, such as *Software Technology*. This indicates that a subset of user studies delves into the technical aspects of software development, implementation, or evaluation, showcasing the intersection of user-focused research with technical domains.

TAKEAWAY (RQ1). The percentage of user studies in HICSS publications is significant at 19%, but the provision of supplementary resources in publicly available external repositories remains low at 1.7%.

Table 1. Distribution of papers with user studies – We show the number of the subclasses of user studies that have functional repositories.

Subclass	Total	Repo?
Users Surveys	472	11
Users Interviews	595	7
Behavioral Studies	42	1

4.2. Technical Papers (RQ2)

To provide answers to RQ2, we discuss technical paper contributions.

Out of 884 papers with a substantial technical component, only 56 (6%) provide supplementary resources in a publicly available external repository. As presented in Fig. 2, the trend of including supplementary resources in an external repository is generally increasing. In 2017, no papers provided an external repository, but this number rose to 6 in 2019, 9 in both 2021 and 2022, and reached 13 in 2023. However, there was a slight decrease to 11 in the year 2024, indicating a fluctuation in the provision of supplementary resources over the years. This trend demonstrates a growing commitment to sharing resources, despite minor variations in the numbers from year to year—a finding which aligns with those by Apruzzese et al. (2023).

Focusing on the contents of the provided repositories, we found out that these usually contain

Table 2. Contents of the repositories for technical papers – We show the most prevalent contents after analysing these repositories.

Content type	Count
Source Code	33
Data	12
Tool	16
Other	9

source code, data, or tools. However, we also added the “other” category for any content that does not fit in any of the aforementioned ones. This is usually related to documentation, knowledge bases, and shared content with user studies. According to our investigations (see Table 2), source code (33, i.e., 59%) is more often shared compared to the other content types. The code is usually related to a certain implementation of an algorithm or a script. This is followed by tool⁵ (16, i.e., 29%) and data (12 i.e., 21%) repositories. Finally, the “other” category was identified 9 (16%) times. Of course, some repositories may have contained more than one of these types of content.

Intriguingly, we found that in 4 (7%) cases, the provided link was not functional, or was not publicly available, or it pointed to an empty repository or a repository was mentioned but with no link (e.g., “code available on GitHub” but without any link). According to this, it can be presumed that in most of cases, the shared repositories are functional and accessible.

Finally, we found that, in general, technical papers represent 16% of the overall number of HICSS publications, which can be considered a significant number for an information systems conference. We also note that, intriguingly, technical papers were more popular than user studies in 2017 (102 vs 98), but this occurrence changes starting from 2018 (88 vs 116): since then, technical papers have always been outnumbered by papers with user studies (e.g., 151 vs 154 in 2024).

TAKEAWAY (RQ2). Only 6% out of 884 technical papers have a link to a publicly available repository. This percentage is superior to user studies, but external repositories are also more valuable for technical papers in the context of reproducibility and usefulness for downstream research (e.g., code reuse).

⁵Tools can be considered as more mature implementations compared to the source-code. For instance, Mclean and Cleveland (2023) provide a tool, whereas only some code snippets and Salminen et al. (2021). Note that neither is better or worse than the other.

4.3. Tracks Distribution (ancillary RQ)

We conclude by considering an ancillary research question: “which *tracks* accept most papers that include a functional link to a corresponding repository?”

To answer such a research question, we show in Table 3, the overall distribution of the papers across the 13 most “stable” tracks⁶ of HICSS during our considered timespan (2017–2024) as well as the percentage of those that provide (or not) an external repository. For instance, for the *Software Technology* track and the *Software, Engineering, Education and Training* track, the overall percentage of papers with an external repository is 9.6% and 6.5%. These results suggest that these tracks are likely to provide more supporting materials to replicate the studies or to gain additional insights into the materials that are not part of the original publication.

In contrast, the *Internet Digital Economy* track and the *Information Technology in Healthcare* track have an overall percentage of papers with an external repository of 1.4% and 1.9%. This indicates that despite the relatively high number of studies conducted in these research areas, there are fewer supporting materials provided. The tracks as the *Internet Digital Economy*, *Collaboration Systems and Technologies*, *Information Technology in Healthcare*, and the *Software, Engineering, Education and Training* track had the highest number of studies, yet relatively lower overall percentage against the repositories.

Looking at the categories of technical papers individually, the tracks *Software, Engineering, Education and Training* as well as *Software Technology* had a percentage of technical papers with an external repository of 11.8% and 11.7%. An example of a recent technical paper with a repository from this particular track is the paper by Pekaric et al. (2024). As for the lowest percentage, the *Information Technology in Healthcare* track had a total of 85 papers, yet only 1.1% contained an external repository. With that said, the overall percentages of the technical papers still ranged from 7.3% to 11.8%. This suggests that the majority of the topics provided a relatively low number of supporting materials—at least compared to papers of highly technical venues, such as those investigated by Olszewski et al. (2023).

The tracks that contained user study papers, on the other hand, resulted in fewer papers providing links to functional external repositories, with the highest having a percentage of 4.3% only (*Software Technology* track). Track *Decisions, Analytics, Services, and Science* had

the lowest percentage of 0.09%. Track *Collaboration Systems and Technologies* was also not far behind at 1.3%. This indicates that the user studies offered fewer supporting materials than the technical papers.

Table 3. Distribution of papers over the tracks – We show the percentage of papers for each “stable” track of HICSS.

Track Name	Technical Papers		User Study	
	Total	Repo?	Total	Repo?
Decis. Analyt. Serv. Scie.	141	10	103	1
Digital Social Media	64	5	83	1
Electric Energy Systems	55	4	1	0
Digital Government	58	3	90	2
Internet Digital Economy	90	1	130	2
Know. Innov. Entr. Syst.	29	2	69	2
Software Technology	140	16	47	2
Location Intelligence	20	0	4	0
Inf. Tech. Social Just.	9	1	8	0
Collab. Syst. Techn.	126	6	151	2
Inf. Tech. Healthcare	85	2	123	2
Soft. Eng. Educ. Train.	17	2	14	0
Org. Syst. Techn.	50	4	174	2

5. Discussion and Lessons Learned

We discuss our findings, outlining limitations and then derive recommendations for future work.

5.1. Threats to validity and Mitigations

We identify two limitations of our analysis.

First, our results stem from a manual analysis carried out after the application of our preliminary filtering script (see §3.2). The script is built on a keyword search of the abstracts of HICSS papers. As such, *it is possible that our results underestimate* the overall number of user studies or technical papers: some papers may very well discuss research that could fall into our definitions (§2.2), but which have not been included in our manual analysis due to their abstract not including our considered search terms. The same applies to repositories: we considered keywords of well-known platforms, but if a paper provided their resource on a custom website, our script may not identify it.⁷

Second, our validity analysis has been carried out qualitatively by two authors, who manually reviewed thousands of papers. It is possible that, during this process, some papers have been assigned to the wrong category, or some links have been overlooked. Due to a lack of ground truth, it is impossible to guarantee “perfect” results. This is why we make our results fully available for scrutiny by fellow researchers: **in our repository, we have included a table showing**

⁶Every year, HICSS may have new tracks. We only consider the tracks that have always appeared at HICSS in our considered timespan.

⁷However, we may still “catch” these cases *if* the text included the other keywords. An example is the work by Milliken et al. (2021).

the breakdown of our analyzed papers, specifying the category, the link to the repository we identified, and the contents we found in such a repository.⁸

A potential workaround to “automatically” address the abovementioned shortcomings is to rely on large language models (LLM). For instance, by submitting our papers to some LLM-based tools⁹ and asking “does the paper entail human-subject research?” or “does the paper include a repository that is specific of the research done in the paper?” it may be possible to derive more accurate results. However, despite their potential, LLMs have explainability issues (Zhao et al., 2024). This is why we preferred to carry out our study “by ourselves,” instead of relying on AI. Nevertheless, given that we publicly release our resources, future work can compare our results with those produced by various LLMs.

5.2. Implications and Disclaimers

To avoid generating harmful misunderstandings, let us clarify the implications of our findings.

First, it is **wrong to conclude** that the overall “lack of external repositories” is a sign that prior work is unreliable or, worse, of poor scientific value. Indeed, there are plenty of reasons why such supplementary material is not provided. For instance, in the case of source code, the owners may not be allowed to share it due to, e.g., industrial secrets. In the case of data, there may be privacy-related reasons.¹⁰ It may even be entirely possible that the paper provides all the necessary elements to replicate the same research—and, hence, there is no need of a third-party repository. Furthermore, additional material for replication can be obtained by contacting the authors. Moreover, for full-fledged systems, it is legitimate to avoid sharing their implementation if the underlying code is not ready for public release: maybe the code is poorly commented, or has no documentation, and publicly releasing such “impractical” code would not be of much use. Finally, HICSS allows to attach “supplementary material” in the submission system: it is possible that the reviewers were provided with extra resources which have been scrutinized, but not released due to any of the abovementioned reasons—thereby further validating the work described in the paper. Therefore, *we refrain from carrying out analyses focusing on individual papers*—as done, e.g., by Königstorfer et al. (2024) and Olszewski et al. (2023)

Nonetheless, it is a fact that only a tiny percentage of papers accepted at HICSS provides publicly

⁸We will update the table if researchers provide factual reasons.

⁹As an example in 2024: <https://www.chatpdf.com/>

¹⁰It may even be that the size of the data is so big that it cannot be included in a single repository due to limitations (Freeman et al., 2022), and the data can hence only be provided upon request.

observable supplementary resources. We believe that the “magnitude” of the research discussed in each paper accepted to HICSS is far greater than the content of the paper itself. Indeed, to quote the overarching message of the “artifact evaluation” process of top-tier security conferences¹¹ “*Your paper is more than just words. Its artifacts extend beyond the document itself: software, hardware, evaluation data and documentation, raw survey results, mechanized proofs, models, test suites, benchmarks, and so on. In some cases, the quality of these artifacts allows for easy extension and reproduction.*” Hence, we endorse the organizers of HICSS’ tracks and respective mini-tracks to promote the release of additional resources, and/or to consider in higher regard those papers that provide such artifacts. Potentially, submitted papers that include artifacts can be marked with specific labels to clearly distinguish them. Finally, we emphasize that, during our manual review, it was not simple to determine if a paper with a dedicated repository actually *had* such a repository, since the link was often hidden among a myriad of other links. Hence, the link to such an artifact could be added to the page of the accepted paper.

5.3. Directions for Future Work

The research discussed in this paper provides a stepping stone for future work—not necessarily pertaining to the field of transparency/reproducibility.

First, our supplementary material includes *the links to all repositories we found* during our analysis. Hence, future work can use such links as a guide to, e.g., build new tools, investigate “why” some authors did (or did not) provide their resources in a publicly available repository, or carry out novel research by relying on the artifacts released by prior publications—thereby further increasing the impact of HICSS’ prior contributions. Alternatively, future work can carry out true reproducibility studies by using the repositories we found and assessing their functionality—as done, e.g., by Olszewski et al. (2023). Such analyses can be beneficial as well as inspirational for those fields of research for which transparency and reproducibility are known to be of crucial importance, such as AI (Gesing et al., 2023).

Another avenue is *expanding our findings*: we release all our resources, so future work can use them to investigate if upcoming editions of HICSS will reflect the same trend as those covered in this paper; or even if editions of HICSS that occurred before our considered timespan (e.g., before 2017) have a different trend. At the same time, it is possible to use our dataset

¹¹A list can be found at: <https://secartifacts.github.io/>.

(which can be recreated by using our scraper §3.1) to carry out orthogonal analyses focused on the overall evolution of the publications of HICSS. For instance, an intriguing phenomenon that we noticed during our manual review is the “different formatting” that HICSS publications tend to have: this can happen either during different editions (e.g., compare Sedlmeir et al. (2021) with Koh et al. (2024)), or even across the same edition of HICSS (e.g., compare Rieger (2018) with Gabbrielli et al. (2018) and Scrivner et al. (2018)).

An intriguing direction is to find ways to improve the research methods used in our work. For instance, one can use large-language models and see if such models can produce similar results to ours in a fraction of the time (it took us weeks to carry out all our manual verifications!). Moreover, it is also possible to improve our script (which is provided in our repository). To this end, we provide two lessons learned from our analysis: (i) keywords can be expanded to include additional synonyms and domain-specific terms; (ii) regular expressions can be utilized for more flexible pattern matching. Alternatively, machine learning approaches, such as topic modeling, can be used to remove hard-coded keyword searches and increase flexibility—potentially allowing to identify papers that we missed, but at the expense of some “false positives”.

Finally, future work can also investigate the papers excluded from our analysis (see §2.2). For instance, some literature reviews can be carried out by means of AI, and may hence include a strong technical component. Whereas some works that are neither technical in nature nor include human-subject research could provide supplementary resources that expand their primary contributions. It is hence intriguing to explore this additional dimension, which would shed more light on HICSS’ history of publications.

6. Conclusions

We have investigated the extent to which HICSS publications within 2017–2024 release additional resources in external repositories. Our analysis, carried out with a mix of scripting and manual reviewing, revealed that only 3% of the 2028 papers that we considered have an external link to an external repository containing supplementary material.

Our efforts can serve as an inspiration for future work: although papers have a limited number of pages, external repositories are not bound by such limits. Such repositories can hence be used to share the dearth of artifacts that are generated in the process of deriving the results of any given research paper. We invite future work to reflect upon our findings: providing supplementary resources not only facilitates

downstream research, but also increases the soundness of each accepted paper—thereby increasing their overall contribution to the state of the art.

References

- Aguinis, H., Cascio, W. F., & Ramani, R. S. (2020). Science’s reproducibility and replicability crisis: International business is not immune. *Research Methods in International Business*.
- Apruzzese, G., Anderson, H. S., Dambra, S., Freeman, D., Pierazzi, F., & Roundy, K. (2023). “Real Attackers Don’t Compute Gradients”: Bridging the Gap Between Adversarial ML Research and Practice. *SaTML*.
- Asprion, P., Hübner, P., & Moriggl, P. (2019). Towards a distributed ledger system for supply chains. *HICSS*.
- Baker, M. (2016). Reproducibility crisis. *Nature*.
- Daneshvar Kakhki, M., Mousavi, R., & Palvia, P. (2021). Evidence quality, transparency, and translucency for replication in information systems survey research. *CAIS*.
- Davis, F. D., Bagozzi, R., & Warshaw, P. (1989). Technology acceptance model. *J. Manag. Sci.*
- Escamilla, E., Klein, M., Cooper, T., Rampin, V., Weigle, M. C., & Nelson, M. L. (2022). The rise of github in scholarly publications. *Int. Conf. Theory and Practice of Digital Libraries*.
- Fišar, M., Greiner, B., Huber, C., Katok, E., Ozkes, A. I., & Collaboration, M. S. R. (2024). Reproducibility in management science. *Management Science*.
- Freeman, N. D., Stubbs, J., & Cardone, R. (2022). Workflow management for scientific research computing with tapis workflows: Architecture and design decisions behind software for research computing pipelines. In *Practice and experience in advanced research computing*.
- Gabbrielli, M., Giallorenzo, S., Lanese, I., & Zingaro, S. P. (2018). A language-based approach for interoperability of iot platforms. *HICSS*.
- Gesing, S., Pierce, M., Marru, S., Zentner, M., Huff, K., Bradley, S., Cleveland, S. B., Brandt, S. R., Ramnath, R., Kee, K., et al. (2023). Science gateways and ai/ml: How can gateway concepts and solutions meet the needs in data science? In *Critical infrastructure-modern approach and new developments*.
- Giudici, M., Abbo, G. A., Crovari, P., & Garzotto, F. (2024). Delivering green persuasion strategies with a conversational agent: A pilot study. *HICSS*.
- Hardwicke, T. E., Wallach, J. D., Kidwell, M. C., Bendixen, T., Crüwell, S., & Ioannidis, J. P. (2020). An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *Royal Society open science*.
- Heine, L., Hörst, F., Nasca, E., Siveke, J., Egger, J., Kim, M., Bahnsen, F. H., & Kleesiek, J. (2024). Lean study host: Towards an automated pipeline for multi-center study hosting. *HICSS*.
- Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*.
- Koh, F., Grosse, K., & Apruzzese, G. (2024). Voices from the frontline: Revealing the ai practitioners’ viewpoint on the european ai act. *HICSS*.
- Königstorfer, F., Haberl, A., Kowald, D., Ross-Hellauer, T., & Thalmann, S. (2024). Black box or open science? a

- study on reproducibility in ai development papers. *HICSS*.
- Majchrzak, T. A., Radianti, J., Fromm, J., & Gau, M. (2022). Towards routinely using virtual reality in higher education. *HICSS*.
- McClean, J., & Cleveland, S. (2023). Design and implementation of web apis for supporting data product visualization and dissemination in science gateways. *HICSS*.
- Milliken, G., Nguyen, S., & Steeves, V. (2021). A behavioral approach to understanding the git experience. *HICSS*.
- Miyakawa, T. (2020). No raw data, no science: Another possible source of the reproducibility crisis.
- Nägele, S., Watzelt, J.-P., & Matthes, F. (2024). Assessing team security maturity in large-scale agile development. *HICSS*.
- Nguyen, A., Gardner, L., & Sheridan, D. (2020). A design methodology for learning analytics information systems: Informing learning analytics development with learning design. *HICSS*.
- Nijssen, J., & Litchfield, A. (2023). Identifying developer engagement in open-source software blockchain projects through factor analysis. *HICSS*.
- Olszewski, D., et al. (2023). "get in researchers; we're measuring reproducibility": A reproducibility study of machine learning papers in tier 1 security conferences. *ACM CCS*.
- Pekarić, I., Felderer, M., & Steinmüller, P. (2021). VULNERLIZER: cross-analysis between vulnerabilities and software libraries. *HICSS*.
- Pekarić, I., Frick, M., Adigun, J. G., Groner, R., Witte, T., Raschke, A., Felderer, M., & Tichy, M. (2024). Streamlining attack tree generation: A fragment-based approach. *HICSS*.
- Plessner, H. E. (2018). Reproducibility vs. replicability: A brief history of a confused terminology. *Frontiers in neuroinformatics*.
- Rieger, C. (2018). Evaluating a graphical model-driven approach to codeless business app development. *HICSS*.
- Rogers, E. M. (2003). *Diffusion of innovations (5th edition)*. Free Press.
- Salminen, J., Jung, S.-g., & Jansen, B. J. (2021). Automatically mapping ad targeting criteria between online ad platforms. *HICSS*.
- Scherer, J., Morais Sanvido, P., Selbach Silveira, M., & Manssour, I. (2022). Integrated visual analysis for brand perception on different social networks. *HICSS*.
- Scrivner, O., Chakilam, V., Sahoo, N., & De Spiegeleire, S. (2018). Topic analysis through streamgraph via shiny application: A social collaborative approach. *HICSS*.
- Sedlmeir, J., Ross, P., Luckow, A., Lockl, J., Miehle, D., & Fridgen, G. (2021). The dlps: A new framework for benchmarking blockchains. *HICSS*.
- Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science advances*.
- Sun, Y., Kim, H. M., Xu, Y., Wang, Y., Kwong, J., Kim, S., Kim, D. O., & McLaughlin, M. (2021). Gps tracking in dementia caregiving: Social norm, perceived usefulness, and behavioral intent to use technology. *HICSS*.
- Taherdoost, H. (2018). A review of technology acceptance and adoption models and theories. *Procedia manufacturing*.
- To, W., & Yu, B. T. (2023). Rise in higher education researchers and academic publications. *Emerald Open Research*.
- Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public opinion quarterly*.
- Ul-Ain, N., Vaia, G., & DeLone, W. (2019). Business intelligence system adoption, utilization and success—a systematic literature review. *HICSS*.
- Yan, B., & Gurkan, N. (2023). It depends on the timing: The ripple effect of ai on team decision-making. *HICSS*.
- Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., Wang, S., Yin, D., & Du, M. (2024). Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology*.