

# Evaluating InterDev: A FAIR Platform for International Development Data\*

Matt Murtagh-White<sup>1</sup>, P. J. Wall<sup>2</sup> and Declan O’Sullivan<sup>3</sup>

<sup>1</sup>CRT-AI, School of Computer Science and Statistics, Trinity College Dublin

<sup>2</sup>ADAPT, Technological University Dublin

<sup>3</sup>ADAPT, School of Computer Science and Statistics, Trinity College Dublin

## Abstract

Over the past twenty years, the application of Randomised Controlled Trials in economics and global development has expanded, offering policymakers and researchers fresh perspectives on effective initiatives. InterDev, an online knowledge discovery platform, enables users to find, discover, and reuse data from evaluations structured according to the ERCT ontology. This study is the first of three iterations evaluating the usability of InterDev through a user study where participants completed 10 tasks, recorded their task completion times, interventions needed, and used the think-aloud protocol. They also filled out the Post-Study System Usability Questionnaire (PSSUQ). Thematic analysis of open-ended responses and recordings, along with quantitative analysis of the PSSUQ, revealed that while users generally find the platform functional, there are significant areas for improvement. Key findings indicate issues with error message clarity and overall user satisfaction, particularly in tasks involving filtering and managing collections. Users highlighted the need for enhanced search capabilities, better guidance and navigation, and more intuitive interface design.

## Keywords

Linked Data, International Development, Randomised Controlled Trials, Knowledge Graph Representation, Data Exploration

## 1. Introduction

In the past two decades, the trend towards evidence-based public policy has catalysed a significant shift in the social sciences, emphasising impact evaluation. Drawing on methodologies from Randomised Controlled Trials (RCTs) in medical research, social scientists and policymakers have embedded evaluation mechanisms into interventionist policies to assess their effectiveness. This research approach has yielded important insights, particularly for public policy in lower-income countries. For instance, studies have shown that childhood exposure to cash transfer programs with conditions tied to health and education can lead to

---

*VOILA 2024: The 9th International Workshop on the Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs co-located with the 23rd International Semantic Web Conference (ISWC 2024), Baltimore, USA, November 11-15, 2024.*

\* Available at <https://interdev.adaptcentre.ie/>

Code available at <https://github.com/mmurtaghw/interDev>

VOILA 2024: The 9th International Workshop on the Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs co-located with the 23rd International Semantic Web Conference (ISWC 2024), Baltimore, USA, November 11-15, 2024.)

✉ [mmurtagh@tcd.ie](mailto:mmurtagh@tcd.ie) (M. Murtagh-White)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

improved educational, mobility, and labour market outcomes in adulthood [1]. Additionally, the duration of exposure to these programs has been linked to increased long-term consumption [2].

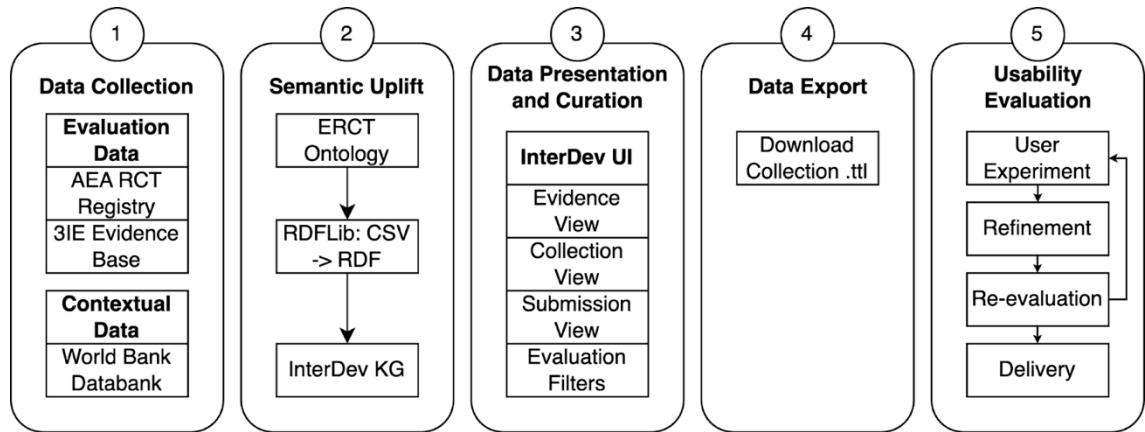
Recently, there has been a growing emphasis on meta-analysis, with researchers seeking to extract broader policy lessons from a deepening pool of evidence [3], [4]. Efforts have been made to create systematic review frameworks to support specific policy areas and address external validity concerns that may arise from conclusions based on single evaluations. Traditional meta-analyses have included both qualitative desk studies that synthesise findings from multiple studies [5], [6] and quantitative approaches that aggregate treatment effects to evaluate the effectiveness of interventions in a particular domain [3], [7].

InterDev, an online knowledge discovery platform, was developed to support this growing need for systematic review frameworks by enabling users to find, discover, and reuse data from evaluations, to make accessing such data Findable, Accessible, Interoperable and Reusable (FAIR) [8]. It builds on previous work on ontology development by providing an interface that allows for the curation of data according to the ERCT ontology framework [9], without the need for knowledge graph expertise that may not be within the remit of non-technical researchers [10]. This follows research which has similarly adapted knowledge graph data for non-technical researchers in health research [11] In this paper, we focus on the user evaluation of InterDev to understand its effectiveness and usability, presenting the first set of results from a planned three round evaluation. Participants were assigned 10 tasks and their task completion times, the number of interventions required, and verbal processes via the think-aloud protocol were recorded by the author and later transcribed. They also completed the Post-Study System Usability Questionnaire (PSSUQ). Through thematic analysis of open-ended responses and recordings, and quantitative analysis of the PSSUQ, we found that users generally navigate the platform well but highlighted the need for additional functionalities, such as enhanced features and improved search capabilities, to maximise its utility.

This paper is structured as follows: Section 2 describes the implementation and methodology of InterDev, detailing data collection, semantic uplift, data presentation, and usability evaluation. Section 3 describes the technical architecture and data integration processes. Section 4 presents the evaluation, including both quantitative results, such as task completion times and PSSUQ scores, and qualitative results from thematic analysis of user feedback. Finally, Section 5 concludes with a summary of findings, discussing strengths and areas for improvement, and outlining future development directions for InterDev.

## **2. Methodology**

The methodology of InterDev can be defined in five key stages, illustrated in figure 1.



**Figure 1:** Overview of Implementation and Methodology.

**Step 1: Data Collection.** In the first phase, evaluation data from various development data sources are gathered along with contextual data from multiple repositories. This comprehensive data collection provides a rich dataset that enables the platform's functionality. Any source of development data where data is structured as evaluations can be integrated.

**Step 2: Semantic Uplift.** The second phase involves the semantic uplift of collected data, where the data is structured according to the ERCT Ontology [9] using tools such as RDFLib [12], allowing for the expression and combination of the underlying data as RDF [13]. This process converts CSV data into RDF (Resource Description Framework) format, facilitating the creation of the InterDev Knowledge Graph (KG). The semantic uplift ensures that data is not only standardized but also enriched with semantic meaning, enhancing the platform's ability to support sophisticated queries and data integration, improving the discoverability and usability of the information.

**Step 3: Data Presentation and Curation.** The third phase focuses on the presentation and curation of data within the InterDev user interface (UI). The platform offers various views, such as Evidence View, Collection View, Submission View, and Evaluation Filters, to help users navigate and interact with the data effectively. This phase is important for transforming raw data into a user-friendly format, enabling users to access, explore, and curate the information they need efficiently, particularly for non-technical researchers unfamiliar with semantic web technology [14].

**Step 4: Data Export.** In the fourth phase, users can export curated data collections in .ttl (Turtle) format. This capability allows users to download and utilize the data outside the platform, facilitating broader dissemination and application of the knowledge discovered through InterDev. Data export is a vital feature for researchers who need to incorporate the data into their analyses or share it with collaborators.

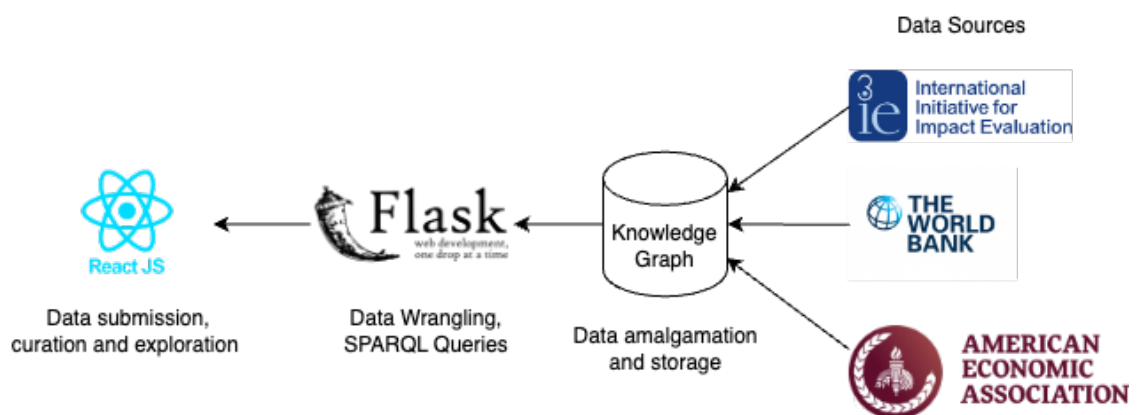
**Phase 5: Usability Evaluation.** The final phase involves a thorough usability evaluation, consisting of user experiments, refinements based on feedback, re-evaluations, and eventual delivery of the improved platform. These evaluations consist of multiple metrics and formats, such as PSSUQ, user interviews and thematic analysis. This phase ensures that the platform meets user needs and expectations, leading to iterative refinements and enhancements based on real user experiences.

Overall, this approach allows for the platform to grow and evolve in response to user feedback to develop a KG powered platform that is shaped by user needs. The KG backend allows for diverse types of data to be integrated into the system, while the incorporation of the ERCT ontology allows the mapping of this data to move towards standardisation. Meanwhile, the development of the InterDev dashboard and frontend allows users who are familiar with international development but do not have technical skills in semantic web technology to take advantage of linked data. As this research develops, the platform is likely to change and adapt in response to each evaluation round.

## 2.1. State of the Art

Existing portals, such as the 3IE Evidence Portal [15] and the American Economic Association’s repository of randomized controlled trials [16], primarily provide high-level overviews and repository functions. InterDev, in contrast, focuses on international development and employs a decentralized, knowledge graph-based approach. This method ensures data consistency and interoperability across diverse datasets. By adopting a single standard for organizing and linking data, InterDev aims to enhance the accessibility and effectiveness of data for policymakers and researchers in the international development sector.

## 3. InterDev Implementation



**Figure 2:** Overview of Framework.

InterDev is designed to provide a knowledge discovery platform aimed at facilitating the integration, curation, and analysis of impact evaluation data within the realm of international development. The architecture of InterDev, shown in Figure 1, is centered around a knowledge graph and an interface developed using React 18.2, with a backend infrastructure supported by Flask 3.0. This setup ensures efficient data discovery and interaction.

### 3.1. Data Collection and Uplift

The data for this study was collected from multiple sources. Data from the International Initiative for Impact Evaluation (3ie) was scraped from their evidence portal, providing

extensive information on the effectiveness of various development interventions. The American Economic Association (AEA) Registry data was obtained through downloadable CSV files, offering detailed records of randomized controlled trials. Additionally, contextual data from the World Bank was sourced from their databank, encompassing a wide range of global development indicators. This multi-source data collection approach underpins the robust knowledge base of InterDev, facilitating thorough analysis and evaluation of development initiatives.

The collected data was uplifted using RDFLib to convert it into RDF (Resource Description Framework) format. This process involved structuring the data according to the ERCT ontology, ensuring consistency and interoperability across different datasets. RDFLib facilitated the transformation of raw data into a standardized format, enabling integration within the knowledge graph.

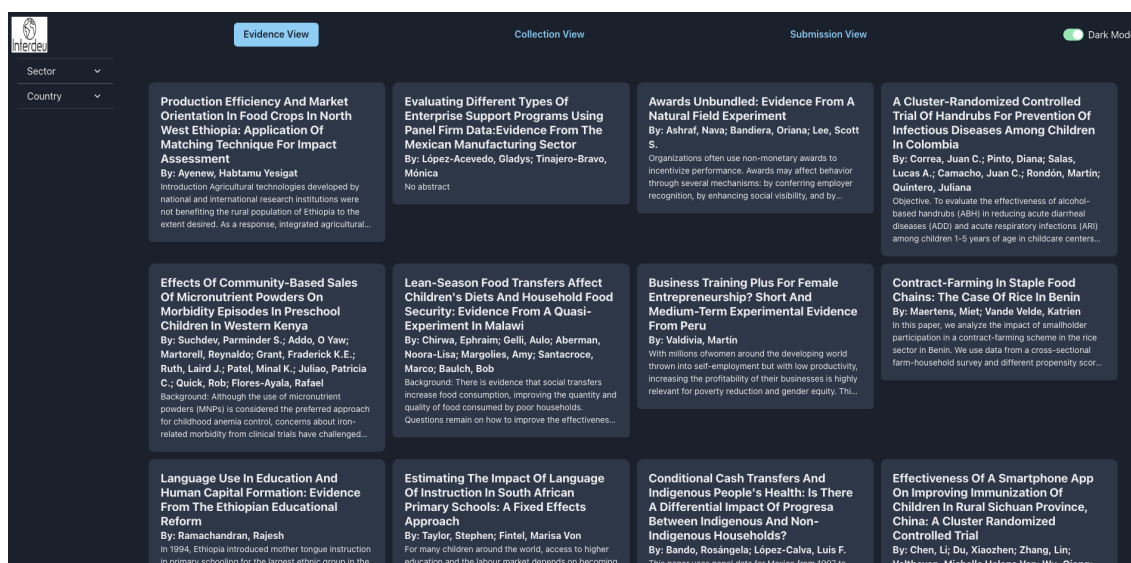


Figure 3: Dashboard View.

### 3.2. Main Dashboard

The main dashboard is the central area for accessing the features of InterDev. It is divided into three sections:

**Navigation Menu:** Located on the left side, this menu provides quick access to key functionalities such as filtering by sector or country.

**Primary Views:** Users can switch between different views (Evidence View, Collection View, Submission View) using buttons at the top.

**Content Area:** The central part displays results of user interactions, such as evidence summaries, collections, or submission forms.

### 3.3. Evidence View

The Evidence View is designed for exploring and searching impact evaluations. Users can refine their searches by filtering results based on criteria such as sector or country. Results are

displayed in a grid format, with each tile representing an evaluation. Tiles provide snapshots including the title, authors, and a brief description. Clicking on a tile gives access to detailed information about the evaluation, including methodology, findings, and related data.

### **3.4. Collection View**

The Collection View allows users to create, manage, and share collections of evaluations for projects or policy decisions. Users can add evaluations from the Evidence View into their collections, view contents, and share or download these collections directly from the platform. This feature facilitates collaboration and the effective utilization of relevant studies.

### **3.5. Submission View**

The Submission View provides an interface for submitting new evaluation data. It guides users through the process to ensure comprehensive and standardized data collection, capturing essential information such as the abstract, authors, title, project details, and evaluation design. This approach adheres to ERCT ontology standards, ensuring submitted data is integrated into the knowledge graph and accessible for future searches and analysis.

## **4. Evaluation**

The first iteration of the InterDev evaluation is described below.

### **4.1. Experimental Design**

The evaluation methodology integrates both qualitative and quantitative approaches. Participants completed ten specific tasks using the InterDev platform, such as searching for evaluations, creating collections, and submitting new data. The study was conducted with five participants, including a mix of PhD researchers and social science researchers, none of whom had prior experience with semantic web technology. The evaluation aimed to assess how effectively users could navigate, interact with, and utilize the platform for their research needs without the requirement of experience in the semantic web. Task completion times were recorded by the observer to measure efficiency. During these tasks, the think-aloud protocol was employed, where participants verbalized their thoughts and actions, providing real-time feedback on their experiences and any difficulties encountered [17].

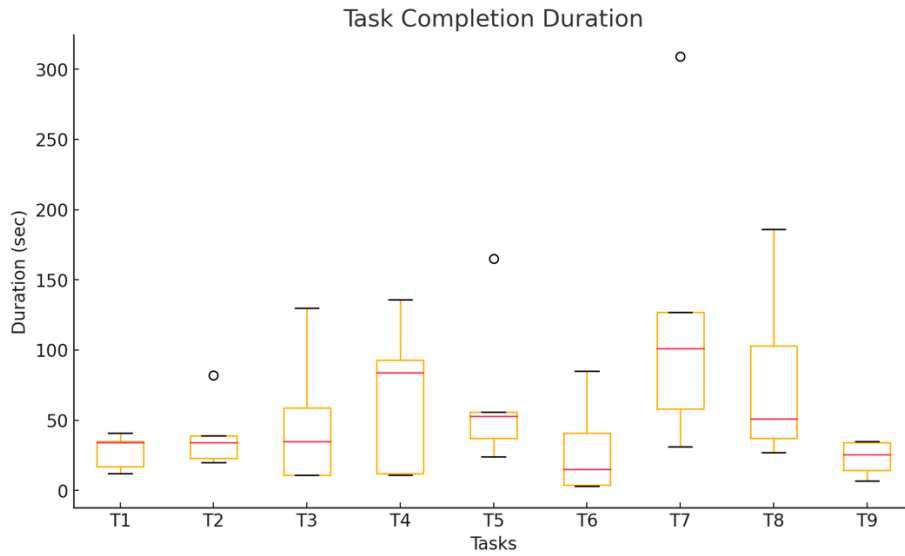
The tasks involved in the evaluation were as follows: selecting “Evidence View” from the navigation bar and waiting for the information to appear (T1), selecting any trial from the evidence view and viewing its associated information (T2), noting the sector of the selected trial (T3), filtering the trials in the evidence view by the noted sector until only trials from that sector appear (T4), adding four trials from this selection to the collection and confirming their presence in the “Collection View” (T5), returning to the evidence view and filtering for both a country and a sector, adding at most four more trials to the collection, and confirming their presence in the “Collection View” (T6), going to the “Collection View,” filtering the collection by any property, and downloading the collection (T7), submitting a new trial with any data in the “Trial Submission” view (T8), finding the submitted evaluation data in the “Evidence View” (T9), and finally, downloading the evaluation data in the .ttl format (T10).

Additionally, instances where participants encountered an issue and required assistance were recorded by the observer to identify potential areas for improvement within the platform. The number of instances of these were recorded by the observer for each task. After completing the tasks, participants filled out the Post-Study System Usability Questionnaire (PSSUQ), which provided quantitative data on their overall satisfaction and the usability of the platform. This is a standardised survey that assessed the evolution of the usability during the development of the system and comprises of 19 questions [18].

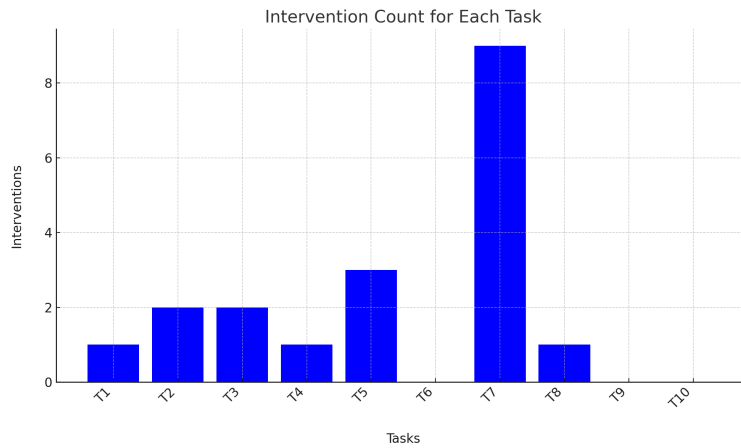
To analyze the data, thematic analysis was conducted on the open-ended responses within the PSSUQ and recordings from the think-aloud protocol, identifying common themes and user feedback. The thematic analysis followed a standardised 6 step process: familiarisation with the data, generation of initial codes, a search for themes, a review of themes, definition and naming of themes and then reporting on findings [19]. Instances of themes were tagged in-text and a script written in Python to count and summarise the instances of these themes across the evaluation data. The PSSUQ results were quantitatively analyzed to assess various aspects of usability, such as ease of use, efficiency, and error handling. This methodology ensures a thorough evaluation of the InterDev platform, combining both user experiences and measurable data to inform future improvements and enhance the platform's usability and effectiveness for researchers and policymakers in international development.

## **4.2. Quantitative Results**

Figure 3 illustrates the box plot of time spent to complete each task. Tasks such as selecting the “Evidence View” from the navigation bar (Task 1), selecting any trial from the evidence view (Task 2), and noting the sector of the trial (Task 3) have low median completion times and minimal variability, indicating that users found these tasks straightforward and easy to complete. However, tasks involving filtering and managing collections presented more challenges. For instance, Task 4, which requires filtering trials for the noted sector, shows moderate median completion time with some variability, suggesting users found the filtering function somewhat challenging. Task 5, which involves adding four trials to the collection, and Task 6, which includes filtering for both a country and a sector, both exhibit higher median completion times and significant variability, indicating these tasks were particularly difficult for users. Other tasks, such as submitting a new trial (Task 8) and finding the submitted evaluation in the evidence view (Task 9), also show higher median completion times and some outliers, reflecting challenges in the submission process and locating submitted evaluations.



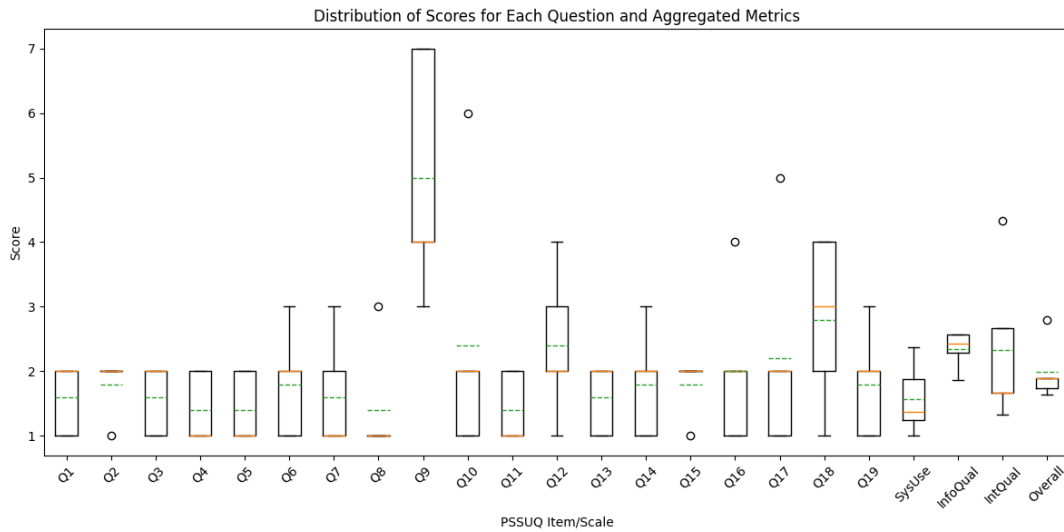
**Figure 4:** Boxplot of time spent to complete each task in the usability evaluation.



**Figure 5:** Bar chart of the number of interventions required for each task.

Figure 4 shows the intervention count for each task, providing further insights into task difficulty. Task 7, which involves filtering the collection by any property and downloading it, had the highest number of interventions, suggesting it was particularly challenging for users. Tasks 2, 3, and 5 had moderate intervention counts, indicating these tasks presented some challenges but were generally manageable. Tasks 1, 4, and 8 had lower intervention counts, suggesting these tasks were relatively straightforward for users. Tasks 6, 9, and 10 had no recorded interventions, indicating that these tasks were the easiest for users to complete independently.





**Figure 6:** Boxplot of PSSUQ scores, based on a Likert 7-point scale where lower values indicate a higher satisfaction. System Usefulness (SysUse), information Quality (InfoQual) and Interface Quality (IntQual) and Overall are aggregated metrics based on questions 1-8, 9-15, 16-18 and 1-19 respectively.

The analysis of the PSSUQ data seen in figure 5 indicates that users generally find the system functional, with lower scores reflecting better usability and satisfaction. However, significant variability in satisfaction levels was observed. Notably, questions related to error messages (Q9) and overall satisfaction (Q19) exhibit higher scores and outliers, suggesting inconsistent user experiences in these areas. This inconsistency underscores the need for targeted improvements in error message clarity and overall system responsiveness. Additionally, the higher median scores for some questions indicate areas where users are less satisfied, highlighting the necessity for comprehensive enhancements in interface design and functionality.

The implications of these findings suggest that while the InterDev platform serves its primary purpose, there is substantial room for improvement. Enhancing error message clarity can significantly reduce user frustration and improve task efficiency, allowing more intuitive interaction with the system.

### 4.3. Qualitative Results

Table 1 summarizes the thematic analysis for the first iteration of InterDev user testing, providing further insights into user feedback. Usability and Interface Design (UID), which encompasses overall design, intuitiveness, and ease of use, had the highest frequency with 19 mentions, indicating that users frequently commented on the visual layout, ease of finding information, and general user experience. Guidance and Navigation (GN) had 13 mentions, highlighting user comments on the clarity of instructions, ease of navigation, and suggestions for improving user guidance, such as better task prompts and visual cues. Functionality and Features (FF) was mentioned 12 times, reflecting feedback related to the platform’s functionalities, including search capabilities, filtering options, and specific features like collection management and submission forms. Efficiency and Performance (EP), with 9

mentions, included observations related to the speed and efficiency of completing tasks, as well as any technical issues or bugs encountered during use.

**Table 1**

Thematic analysis summary for the first iteration of InterDev user testing.

Theme	Code Description	Frequency
Usability and Interface Design (UID)	Overall design, intuitiveness, and ease of use of the platform's interface. This includes feedback on visual layout, ease of finding information, and general user experience.	19
Guidance and Navigation (GN)	Comments on the clarity of instructions, ease of navigation, and suggestions for improving user guidance, such as better task prompts and visual cues.	13
Functionality and Features (FF)	Feedback related to the platform's functionalities, such as search capabilities, filtering options, and specific features like collection management and submission forms.	12
Efficiency and Performance (EP)	Observations related to the speed and efficiency of completing tasks, as well as any technical issues or bugs encountered during use.	9

## 5. Conclusion

The initial evaluation of InterDev demonstrates its potential to enhance data discovery and usability for researchers and policymakers in international development. While users found the platform generally functional, significant improvements are needed, particularly in filtering, error message clarity, search capabilities, and overall interface design. Quantitative and qualitative feedback from our user study highlighted key areas for enhancement, such as better guidance, improved navigation, and more intuitive features. These insights will guide the iterative refinement of InterDev to better meet user needs. While InterDev shows promise, further continuous user-centered development is required. Future iterations will address identified challenges, further refine user needs, and aim to improve the user experience and maximize the platform's utility in making international development data more accessible and actionable.

## Acknowledgements

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106\_P2 at the ADAPT SFI Research Centre at Trinity College

Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme.

## References

- [1] D. de Walque, L. Fernald, P. Gertler, and M. Hidrobo, “Cash transfers and child and adolescent development,” 2017.
- [2] S. W. Parker and T. Vogl, “Do conditional cash transfers improve economic outcomes in the next generation? Evidence from Mexico,” National Bureau of Economic Research, 2018.
- [3] S. Baird, F. H. Ferreira, B. Özler, and M. Woolcock, “Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: a systematic review,” *Campbell Syst. Rev.*, vol. 9, no. 1, pp. 1–124, 2013.
- [4] F. Bastagli *et al.*, “Cash transfers: what does the evidence say,” *Rigorous Rev. Programme Impact Role Des. Implement. Featur. Lond. ODI*, vol. 1, no. 7, 2016.
- [5] H. Waddington *et al.*, “How to do a good systematic review of effects in international development: a tool kit,” *J. Dev. Eff.*, vol. 4, no. 3, pp. 359–387, 2012.
- [6] R. T. Edwards, J. M. Charles, and H. Lloyd-Williams, “Public health economics: a systematic review of guidance for the economic evaluation of public health interventions and discussion of key methodological issues,” *BMC Public Health*, vol. 13, no. 1, pp. 1–13, 2013.
- [7] J. Peters, J. Langbein, and G. Roberts, “Policy evaluation, randomized controlled trials, and external validity—A systematic review,” *Econ. Lett.*, vol. 147, pp. 51–54, 2016.
- [8] M. D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and stewardship,” *Sci. Data*, vol. 3, no. 1, pp. 1–9, 2016.
- [9] M. Murtagh-White, “ERCT: An Ontology for Describing Randomised Controlled Trials in the Social Sciences,” 2021.
- [10] K. Smith-Yoshimura, “Analysis of 2018 international linked data survey for implementers,” *Code4Lib J.*, no. 42, 2018.
- [11] A. Navarro-Gallinad, F. Orlandi, and D. O’Sullivan, “Enhancing rare disease research with semantic integration of environmental and health data,” presented at the Proceedings of the 10th International Joint Conference on Knowledge Graphs, 2021, pp. 19–27.
- [12] RDFLib, “RDFlib.” [Online]. Available: <https://pypi.org/project/rdflib/>
- [13] D. Brickley, R. V. Guha, and B. McBride, “RDF Schema 1.1,” *W3C Recomm.*, vol. 25, pp. 2004–2014, 2014.
- [14] L. Rietveld and R. Hoekstra, “The YASGUI family of SPARQL clients 1,” *Semantic Web*, vol. 8, no. 3, pp. 373–383, 2017.
- [15] International Initiative for Impact Evaluation, “3ie Development Evidence Portal.” Accessed: Apr. 04, 2022. [Online]. Available: <https://www.3ieimpact.org/evidence-hub>
- [16] American Economic Association, “Trial Data Access,” Trial Data Access. Accessed: Jul. 13, 2021. [Online]. Available: <https://www.socialscienceregistry.org/site/data>
- [17] T. Boren and J. Ramey, “Thinking aloud: Reconciling theory and practice,” *IEEE Trans. Prof. Commun.*, vol. 43, no. 3, pp. 261–278, 2000.
- [18] J. R. Lewis, “Psychometric evaluation of the PSSUQ using data from five years of usability studies,” *Int. J. Hum.-Comput. Interact.*, vol. 14, no. 3–4, pp. 463–488, 2002.

- [19] L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules, "Thematic analysis: Striving to meet the trustworthiness criteria," *Int. J. Qual. Methods*, vol. 16, no. 1, p. 1609406917733847, 2017.