Int. j. adv. multidisc. res. stud. 2026; 6(6):799-804

Received: 29-09-2026
Accepted: 09-11-2026

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Knowledge Discovery and Researcher Recommendation in Digital Research Ecosystems

[1] Hellen B Tembo, [2] Abi Nithisha
[1] DMI–ST. Eugene University, Chibombo Campus, Zambia
[2] Guide, DMI–ST. Eugene University, Chibombo Campus, Zambia

Corresponding Author: **Hellen B Tembo**

Abstract
Research activities are frequently distributed across separate digital services for scholarly literature, researcher identity, research data, projects, funding and institutional information. This fragmentation can make it difficult to discover relevant knowledge, identify potential collaborators and translate research ideas into actionable projects. This paper presents the design, implementation and validation of a web-based National Open Science and Research Collaboration Hub developed to bring these activities into an integrated digital research environment. The application is implemented with Python and the Django web framework and provides researcher registration and profiles, publication and dataset management, project workspaces, institutional discovery, funding opportunities, collaboration requests, direct messaging, notifications, research analytics and artificial intelligence (AI)-assisted services. Its researcher recommendation component constructs weighted researcher representations from profile attributes and scholarly records and ranks candidate researchers using term frequency-inverse document frequency (TF-IDF) weighting and cosine similarity. The platform also incorporates a research assistant that combines intent recognition, retrieval from hub records and external scholarly sources, conversational context and an optional language-model service. Additional research-intelligence functions support topic analysis, research-gap exploration, proposal development and project-impact monitoring. Functional, integration and AI-feature testing were completed on the working application. The results establish that the implemented platform provides an operational environment for connected research discovery, researcher matching and collaboration support.

Keywords: Open Science, Knowledge Discovery, Researcher Recommendation, Research Collaboration, TF-IDF, Cosine Similarity, Django, Artificial Intelligence

1. Introduction

Digital technologies have changed how research is conducted, communicated and reused. Researchers increasingly depend on online systems to locate scholarly publications, identify experts, manage research outputs, discover funding and establish collaborative relationships. These services, however, are commonly distributed across independent repositories, publisher platforms, institutional websites, researcher identity systems and funding portals. Users may therefore need to search several services and manually connect information that is related in subject but stored in different places.

Open science provides a framework for making scientific knowledge more accessible, transparent and reusable. UNESCO (2021) [12] identifies open scientific knowledge, open science infrastructures and engagement with societal actors as its key dimensions, and the FAIR principles (Wilkinson *et al*., 2016) [13] emphasize that research data should be findable, accessible, interoperable and reusable. Open scholarly metadata services such as OpenAlex (Priem *et al*., 2022) [8], Crossref (Hendricks *et al*., 2020) [3] and Semantic Scholar (Kinney *et al*., 2023) [4] show the value of interoperable metadata, while studies of data sharing and open access (Zuiderwijk *et al*., 2020; Tennant *et al*., 2016; Fecher & Friesike, 2014) [14, 11, 2] support platforms that connect publications with datasets and projects.

Persistent identifiers such as ORCID (ORCID, n.d.) [7] help distinguish researchers and connect their work, and the study of co-authorship networks (Newman, 2001) [6] provides a basis for making collaborations easier to explore. Recommender systems help users navigate large information spaces (Resnick & Varian, 1997) [9], and content-based, collaborative and hybrid approaches can draw on multiple information sources (Adomavicius & Tuzhilin, 2005) [1]. A content-based approach is

appropriate where explicit profile and scholarly-record information exists but user interaction histories are not yet extensive.

Despite this, these functions are usually met through separate services. Relevant publications, datasets, funding calls and researcher profiles are kept in different systems, which increases the effort needed to find related information and to identify collaborators whose interests, publications and projects align with a researcher's own. A connected platform is therefore required to organize research information, support cross-category discovery and provide explainable recommendations.

The aim of this study was to design, implement and validate an integrated digital research platform that supports knowledge discovery, researcher recommendation and research collaboration. The platform combines locally managed research records with selected external scholarly services, and adds a conversational research assistant, personalized recommendations and research-intelligence tools. Its main contribution is a working platform that uses transparent, content-based researcher matching over several profile and research-record fields, and connects discovery with research planning and project monitoring.

2. Materials and Methods

2.1 Materials and Technologies

The platform was developed iteratively through requirements identification, system and interface design, module development, integration and testing. Each module was checked before being connected to the wider application. The technologies used are summarized in Table 1.

Table 1: Components and technologies of the platform

Component	Technology or method	Purpose
Front end	Django templates, HTML, CSS and JavaScript	Provides the user interface
Back end	Python and Django	Implements application logic, authentication and access control
Database	SQLite through Django ORM	Stores user, research and collaboration records
API	Django REST Framework	Supports API functionality
External services	OpenAlex, ROR, DataCite, Zenodo, OpenAIRE Graph and NIH RePORTER	Literature, institution, dataset and project discovery
Recommendation	TF-IDF and cosine similarity	Ranks researchers and resources
AI assistant	Intent recognition, retrieval and optional language model	Answers research queries
Access control	Visibility rules and private profile fields	Restricts access to records and contacts

2.2 System Architecture

The application follows a layered web architecture, shown in Fig 1.1. The presentation layer provides the registration, dashboard, profile, discovery, collaboration and research-management interfaces. The application layer, written in Python with Django views, models, forms and services, handles authentication, profile management, research records, recommendation, search, collaboration, funding, notifications and research-intelligence operations. The data layer uses Django's object-relational mapping (ORM) with SQLite.

External information is obtained through dedicated client services that normalize responses from OpenAlex and the Research Organization Registry (ROR) for literature and institutions, DataCite and Zenodo for datasets, and OpenAIRE Graph and the United States National Institutes of Health (NIH) RePORTER for research projects. Requests use caching, timeouts and normalized result structures so that the interface can process records consistently.

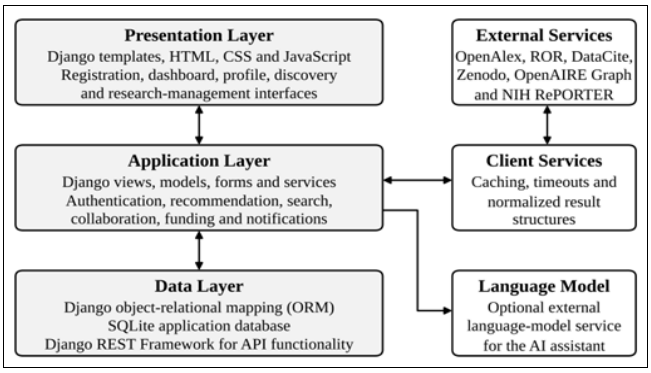


Fig 1.1: Layered Architecture of the Research Collaboration Hub

Explanation of the architecture

1. **Presentation layer:** Provides the screens for registration, discovery, collaboration and research management.
2. **Application layer:** Uses Python and Django to process requests, apply access rules and run recommendation and research-intelligence logic.
3. **Data layer:** Stores user accounts, profiles, publications, datasets, projects and messages through the Django ORM.
4. **Client services:** Retrieve, cache and normalize records from external scholarly and research-data providers.
5. **AI services:** Support the research assistant, using an optional language-model service with a built-in retrieval pathway as fallback.

2.3 Researcher Recommendation Method

A researcher's representation is built from profile fields (research interests, field of study, department and biography) and scholarly evidence (publication titles, keywords and abstracts, and project titles and research areas). These fields are weighted: research interests and publication keywords receive greater weights than general biographical information, and a researcher's own projects contribute more strongly than projects in which the researcher is a collaborator.

Each researcher is represented as a sparse term frequency–inverse document frequency (TF-IDF) vector (Salton & Buckley, 1988) [10]. Text is tokenized, stop words are removed, light stemming is applied, and adjacent words are also represented so that multi-word research topics contribute to matching. For a term t in researcher document d, sub-linear term frequency and smoothed inverse document frequency are used:

$$tf(t,d) = ln(1 + weighted\ count(t,d)) \quad (1)$$

$$idf(t) = ln((1 + N) \div (1 + df(t))) + 1 \quad (2)$$

Where N is the number of researcher documents in the indexed collection and df(t) is the number of documents containing term t. The term weight is tf(t,d) multiplied by idf(t), after which each vector is L2-normalized. The similarity between the signed-in researcher's vector A and another researcher's vector B is then the cosine similarity (Manning *et al*., 2008) [5]:

$$cosine(A,B) = \Sigma t\ (At \times Bt) \quad (3)$$

Candidate researchers are ranked in descending order of similarity, the signed-in user is excluded, and a minimum similarity threshold is applied. The interface displays shared terms and a short explanation so that the basis of each recommendation is understandable. The behavior of the two weighting components is illustrated in Fig 1.2, using an assumed collection of N = 100 researcher documents for the inverse document-frequency curve.

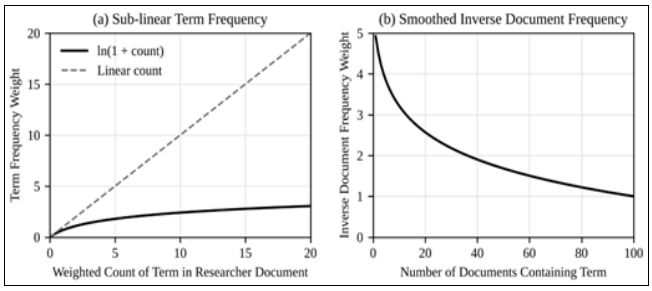


Fig 1.2: Term-Frequency and Inverse Document-Frequency Weighting Behavior

The recommendation module also builds a weighted interest profile from a user's research interests, publication keywords, project research areas and department, expands related terms using the platform's topic vocabulary, and compares the profile with the titles, keywords, areas and descriptions of publications, datasets, projects and funding opportunities. When a profile has insufficient information, a cold-start pathway presents recent projects and available funding calls rather than an empty page.

2.4 AI Research Assistant and Research Intelligence

The research assistant follows a retrieval-and-response workflow. A submitted question is processed to identify its intent and topic terms; relevant hub records (publications, datasets, projects, funding, researchers and institutions) are retrieved and, where relevant, external scholarly services are queried. The assistant uses the retrieved evidence, the user's profile and recent conversation history to formulate a response. When an external language-model service is configured it generates a contextual response; otherwise the built-in retrieval and response pathway is used.

The research-intelligence component analyzes indexed publications, projects and datasets to identify topic distributions, trends and evidence-backed research-gap signals, with findings linked to the records used so that users can inspect the evidence. These findings reflect the indexed material only and are not an assertion that research does not exist elsewhere. The proposal assistant uses related publications and projects to help structure a proposal, and project-impact monitoring records milestones, progress reports, outputs and outcomes.

2.5 Data Privacy and Access Control

Access to user-specific functions requires authentication, and publication, dataset and project records support visibility rules that govern which records a viewer can see. Researcher profile fields can be designated as private, with contact details private by default. These controls are integrated into the application workflows to support responsible access to researcher information and shared research outputs.

3. Results and Discussion

3.1 Results

The implementation resulted in a working web-based platform. The account module supports registration, authentication and researcher-profile maintenance, storing name, affiliation, department, country, province, field of study, ORCID identifier, research interests and biography; ORCID values are normalized and validated before being accepted. The publication, dataset and project modules store the title, description, keywords, category, file and status information of each record, and the Explore interface brings together discovery pages for publications, datasets, projects, researchers and institutions, using visibility-aware queries for local records.

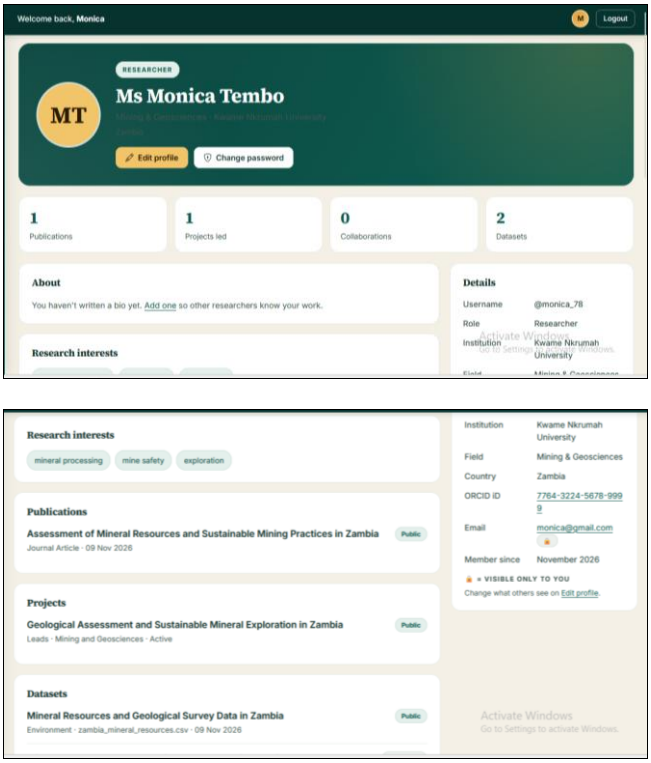
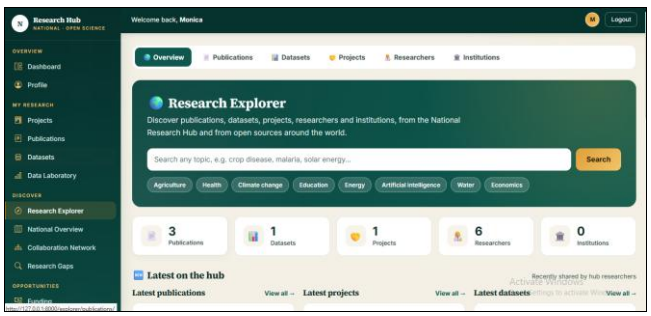


Fig 1.3: Researcher profile



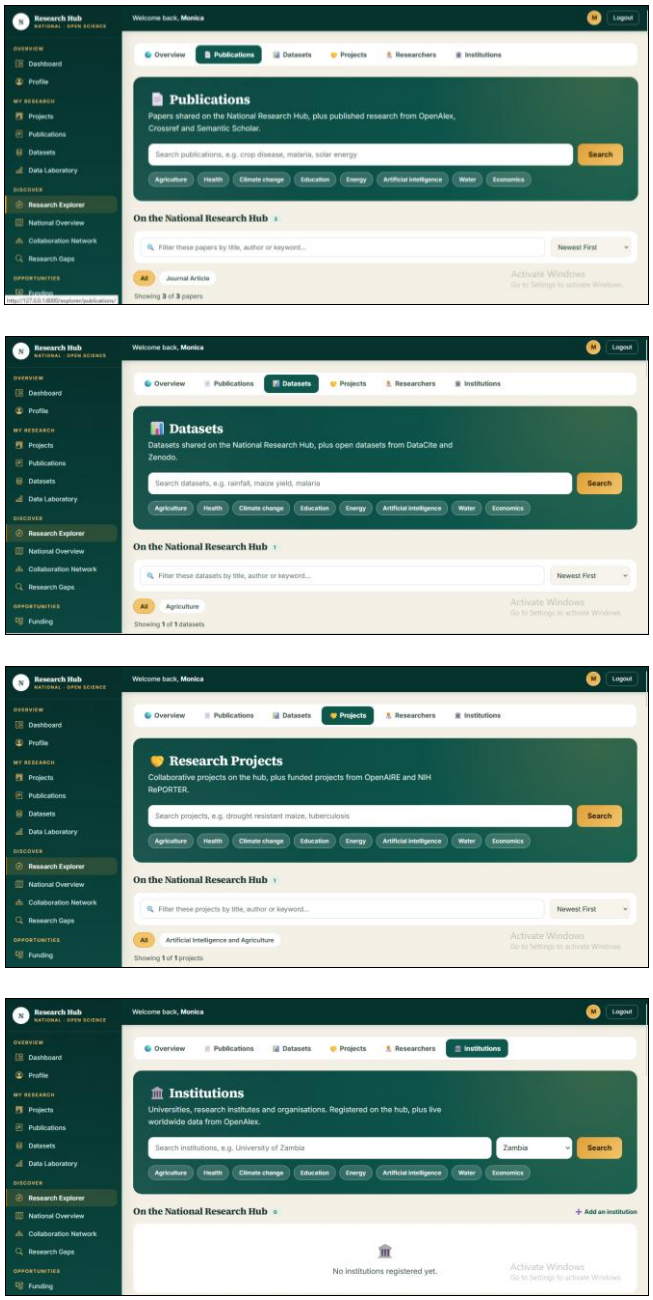


Fig 1.4: Explore interface for discovery of research records

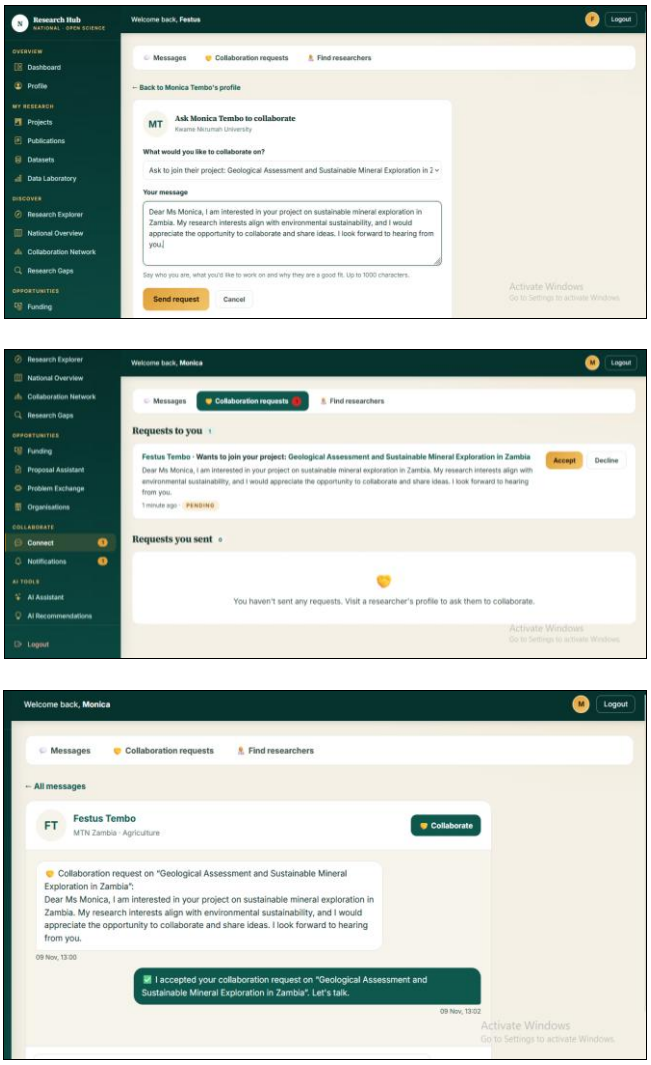
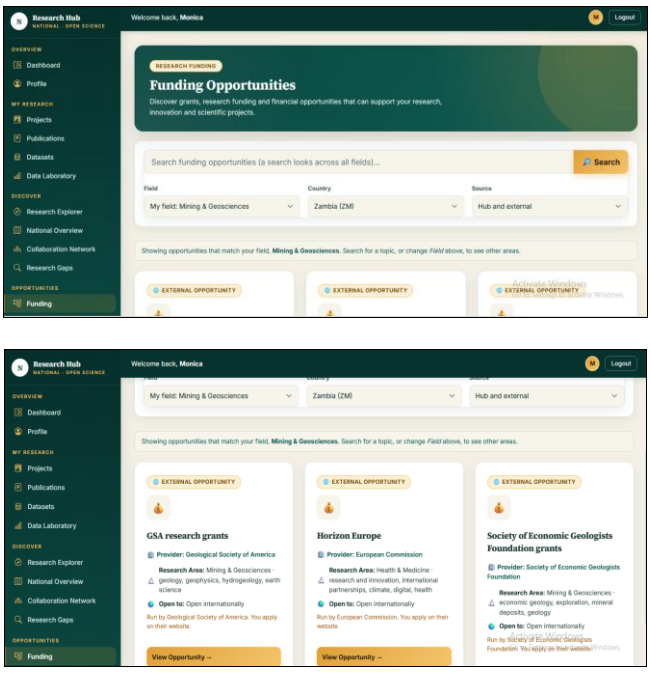
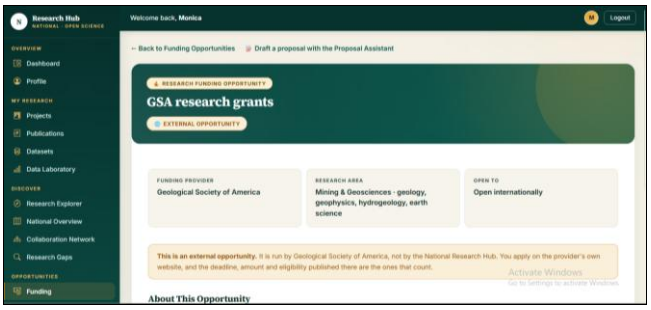


Fig 1.5: Collaboration requests and messaging module



The collaboration module lets a researcher send a request to join or invite a researcher to a project, or to propose general collaboration, with each request recording sender, recipient, message, status and response time. Direct messaging records timestamps and read status, and notifications report relevant platform activity. The funding module presents opportunities posted within the hub and from external sources, including provider, eligibility, deadline, application link and research area; external opportunities direct users to the provider's own application information. The AI interface offers a conversational page with stored chat history and suggested queries, and the recommendation page groups funding calls, projects, researchers, publications and datasets with matching reasons.



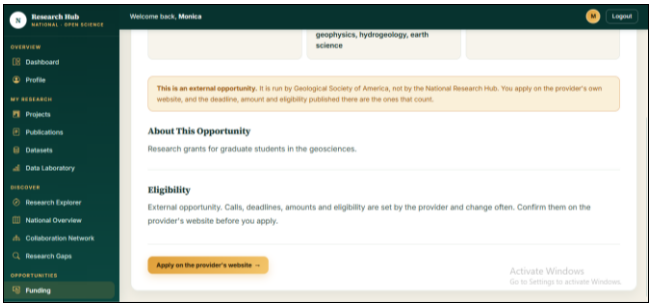


Fig 1.6: Funding opportunities page

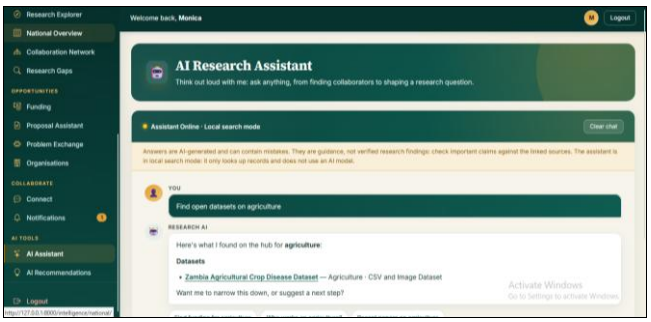
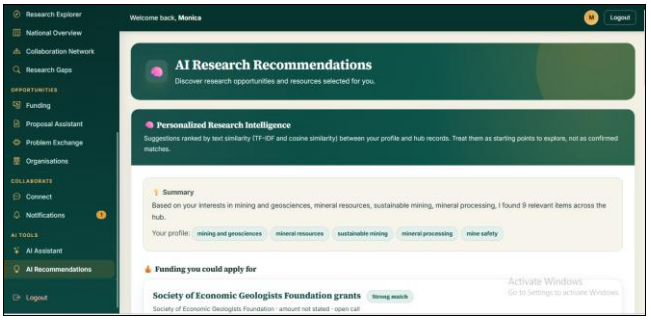


Fig 1.7: AI research assistant interface



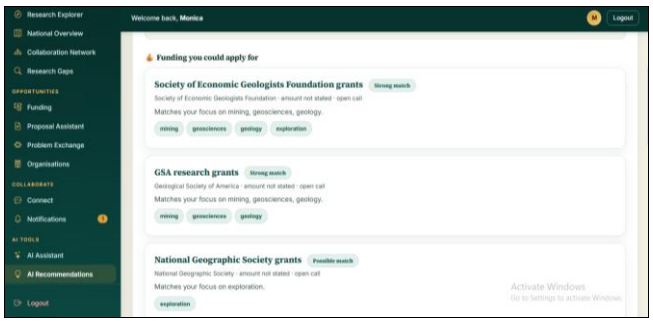


Fig 1.8: Recommendation page with matching reasons

The completed application was tested across its principal user journeys and modules. The results are summarized in Table 2: all tested areas operated as intended.

Table 2: Functional test coverage and observed outcome

Test area	Validation focus	Outcome
Registration and authentication	Account creation, sign-in and authenticated access.	Passed
Researcher profiles	Storage and display of identity, ORCID and academic profile information.	Passed
Research records	Publication, dataset and project management and discovery.	Passed
External discovery	Access to literature, dataset, project and institution discovery interfaces.	Passed
Collaboration and messaging	Collaboration requests, responses and direct communication.	Passed
Funding and notifications	Presentation of opportunities and platform updates.	Passed
AI research assistant	Question handling, retrieval and response generation.	Passed
Researcher recommendations	Similarity-based ranking and display of relevant researcher profiles.	Passed
Research intelligence	Topic exploration, gap analysis and related research-support pages.	Passed
Proposal and impact tools	Proposal structuring, milestones, progress, outputs and outcomes.	Passed
Navigation and integration	Movement between the main modules and integrated user journey.	Passed

Integration testing covered the interaction of profile information with recommendations, the presentation of local records, navigation between discovery categories, collaboration and messaging workflows, funding and notification pages, and access to the AI and research-intelligence facilities. The external-service clients return normalized records, with caching and error handling built into the integration layer.

The recommendation method uses more than the research-interest field alone: profiles are enriched with publication titles, keywords and abstracts, project titles and research areas, and selected profile attributes. TF-IDF weighting reduces the influence of terms common to many researchers and increases the contribution of distinctive terms, and cosine similarity gives a normalized basis for ranking. Shared terms are displayed with each recommendation, giving users an interpretable indication of why a researcher was suggested. The same weighted profile supports personalized discovery of projects, publications, datasets and funding opportunities.

3.2 Discussion

The implemented platform shows how open scholarly metadata, researcher identity information, content-based recommendation and collaboration tools can be combined in a single web application. Its main practical value lies in reducing the separation between research discovery and research interaction: a researcher can manage outputs, explore related work, identify potential collaborators, review funding information and use AI-assisted support without treating each activity as an isolated workflow. The recommendation method is grounded in explicit profile and research-record information, and the assistant combines retrieval with response generation to keep answers connected to available evidence.

Integrated discovery: Local and external records are searched and presented through common structures.
Explainable matching: Shared terms show why a researcher or resource was recommended.
Collaboration support: Requests, messaging and notifications connect discovery with interaction.
AI-assisted support: The assistant retrieves hub and external evidence before generating responses.
Research intelligence: Topic, trend and research-gap findings link back to the supporting records and reflect only the indexed material.
Responsible access: Visibility rules and private profile fields control what is shown to other users.

4. Conclusion

This study designed, implemented and tested a web-based National Open Science and Research Collaboration Hub to support knowledge discovery, researcher recommendation and collaboration in a unified environment. The platform combines researcher profiles, publications, datasets, projects, institutions, funding opportunities, collaboration requests, messaging and analytics with an AI research assistant, a TF-IDF and cosine-similarity researcher recommender, research-gap intelligence, proposal assistance and project-impact monitoring.

Functional and integration testing confirmed the operation of the principal modules within the project environment, and the recommender provides ranked suggestions from multiple profile and scholarly-record fields, with shared terms to support interpretation.

Future work can extend the range of participating repositories and data providers, increase the size of the indexed corpus, and evaluate recommendation quality, diversity and user satisfaction with larger and more diverse researcher populations. The usefulness of research-gap findings and the effectiveness of AI-assisted proposal development should also be examined.

5. References

1. Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering. 2005; 17(6):734-749. Doi: https://doi.org/10.1109/TKDE.2005.99
2. Fecher B, Friesike S. Open science: One term, five schools of thought. In Opening Science. Springer, 2014, 17-47. Doi: https://doi.org/10.1007/978-3-319-00026-8_2
3. Hendricks G, Tkaczyk D, Lin J, Feeney K. Crossref: The sustainable source of community-owned scholarly metadata, 2020. Doi: https://doi.org/10.1629/uksg.491
4. Kinney R, Anastasiades C, Authur R, *et al*. The Semantic Scholar Open Research Corpus, 2023. Doi: https://arxiv.org/abs/2301.10140
5. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge University Press, 2008. https://nlp.stanford.edu/IR-book/
6. Newman MEJ. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences. 2001; 98(2):404-409. Doi: https://doi.org/10.1073/pnas.021544898
7. ORCID. What is an ORCID iD and how do I use it? ORCID Support, n.d. https://support.orcid.org/hc/en-us/articles/360006897454-What-is-my-ORCID-iD
8. Priem J, Piwowar H, Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, and concepts, 2022. https://arxiv.org/abs/2205.01833
9. Resnick P, Varian HR. Recommender systems. Communications of the ACM. 1997; 40(3):56-58. Doi: https://doi.org/10.1145/245108.245121
10. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Information Processing & Management. 1988; 24(5):513-523. Doi: https://doi.org/10.1016/0306-4573(88)90021-0
11. Tennant JP, Waldner F, Jacques DC, *et al*. The academic, economic and societal impacts of open access: An evidence-based review. F1000Research. 2016; 5:632. Doi: https://doi.org/10.12688/f1000research.8460.3
12. UNESCO. UNESCO recommendation on open science. United Nations Educational, Scientific and Cultural Organization, 2021. https://unesdoc.unesco.org/ark:/48223/pf0000379949
13. Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al*. The FAIR guiding principles for scientific data management and stewardship. Scientific Data. 2016; 3:160018. Doi: https://doi.org/10.1038/sdata.2016.18
14. Zuiderwijk A, Shinde R, Jeng W. What drives and inhibits researchers to share and use open research data? Science and Public Policy, 2020.