Crimson Publishers
Wings to the Research

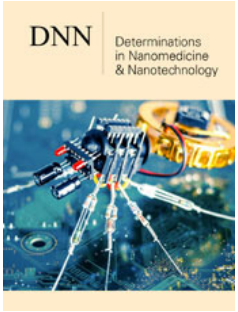
Research Article

Comparative Observational Study of Google Scholar and ProQuest Databases for Doctoral Dissertations on Nanotechnology Applications in Medicine: Implications for Library Science and Archival Practice

Gregory K Tharp*

Independent Researcher, USA

ISSN : 2832-4439



***Corresponding author:** Gregory K Tharp, MLS, Adv Cert, MCAC, CPhT-Adv, USA

Submission: October 16, 2026
Published: November 25, 2026

Volume 3 - Issue 5

How to cite this article: Gregory K Tharp*. Comparative Observational Study of Google Scholar and ProQuest Databases for Doctoral Dissertations on Nanotechnology Applications in Medicine: Implications for Library Science and Archival Practice. Determ in Nanomed & Nanotech. 3(5). DNN. 000573. 2026.
DOI: 10.31031/DNN.2026.03.000573

Copyright© Gregory K Tharp, This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract

Background: Nanotechnology applications in medicine-encompassing targeted drug delivery systems, diagnostic nano sensors, and regenerative scaffolds-represent a rapidly expanding, highly interdisciplinary domain. A significant portion of novel research in nanomedicine originates in graduate dissertations. However, discovering and managing this gray literature depends heavily on the indexing depth, search capabilities, and preservation infrastructure of academic search engines, institutional repositories, and commercial databases.

Objective: This observational study systematically evaluates and compares Google Scholar (GS) and ProQuest Dissertations & Theses Global (PQDT) in indexing and retrieving doctoral dissertations on medical nanotechnology, while examining direct applications for library science, cataloging, Institutional Repository (IR) management, and digital preservation.

Methods: An observational comparative retrieval design was executed using standard Boolean search strings across both platforms. Queries focused on core nanomedicine domains, including drug delivery, targeted cancer therapeutics, diagnostic imaging, and nanotoxicity. Search results were evaluated across bibliographic metrics (recall, precision, metadata accuracy, access) alongside archival and Library Information Science (LIS) parameters (controlled vocabulary, OAI-PMH harvesting, and digital preservation stability).

Results: Google Scholar demonstrated broader general recall and superior full-text keyword discovery, indexing interdisciplinary nanomedicine dissertations housed across institutional repositories. However, Google Scholar exhibited lower precision and frequent metadata parsing errors due to automated web crawling. ProQuest PQDT demonstrated near-perfect metadata accuracy, superior search field filtering (e.g., subject classifications, advisor name, degree-granting institution), and controlled embargo handling, though full-text access for non-open-access items remained gated behind institutional subscriptions.

Conclusion: While Google Scholar serves as an effective high-recall discovery engine for interdisciplinary nanomedicine topics, ProQuest PQDT remains the definitive resource for structured, high-precision searches, library cataloging integration, and long-term archival curation. Information professionals should employ a hybrid discovery and repository strategy to optimize access, bibliographic control, and digital preservation.

Introduction

Nanotechnology applications in medicine, frequently termed nanomedicine, leverage the unique chemical, physical, and biological properties of materials engineered at the nanoscale (1-100nm) to diagnose and treat biological systems [1]. Key developments in nanomedicine include nanoscale drug carriers that enhance therapeutic delivery while minimizing systemic

toxicity, biocompatible diagnostic contrast agents, and advanced materials for cellular engineering [1]. Because nanomedicine sits at the confluence of biomedical engineering, chemistry, pharmacology, materials science, and clinical oncology, early-stage breakthroughs are frequently documented first in doctoral dissertations before being condensed into peer-reviewed journal articles [2]. Doctoral dissertations represent a fundamental class of scholarly gray literature [3]. However, discovering, organizing, and preserving these complex works presents distinct challenges for academic libraries and archives. Electronic Theses and Dissertations (ETDs) are distributed across localized Institutional Repositories (IRs), national libraries, and commercial databases.

Two primary platforms dominate electronic thesis and dissertation discovery and management:

1. Google Scholar (GS): A freely accessible, web-crawling academic search engine that indexes multidisciplinary literature, including institutional repository records and commercial repository feeds [4].

2. ProQuest Dissertations & Theses Global (PQDT): A specialized subscription database providing curated, bibliographically rich indexing and archiving of master's theses and doctoral dissertations from over 3,100 universities worldwide.

Understanding the comparative strengths, recall efficiency, metadata fidelity, and archival stability of Google Scholar versus ProQuest [4] PQDT is essential for biomedical researchers, academic librarians, metadata catalogers, and digital archivists. This observational study analyzes the search performance and indexing characteristics of both platforms when querying doctoral dissertations on nanotechnology applications in medicine, concluding with specific strategic applications for library science and archival workflows.

Methodology

Research design

This study employed a cross-sectional observational retrieval methodology to evaluate database coverage, search recall, precision, metadata structure, and archival integration across Google Scholar and ProQuest Dissertations & Theses Global.

Search strategy & queries

Standardized keyword sets were derived from key sub-disciplines within medical nanotechnology, including nanotoxicity, targeted drug delivery, and diagnostic imaging [4]. The primary query string was formulated as:

("nanotechnology" OR "nanomedicine" OR "nanoparticle*") AND ("medicine" OR "drug delivery" OR "diagnostics" OR "theranostics") AND ("doctoral dissertation" OR "Ph.D. thesis")

1. Google Scholar Parameters: Executed using advanced search operators restricted to document types containing dissertation metadata, evaluated across the first 200 returned hits.

2. ProQuest PQDT Parameters: Executed using the advanced search interface, filtered explicitly by Document Type: Dissertation/Thesis and Degree Level: Doctoral.

Evaluation metrics

Each platform was evaluated across six observational parameters:

1. Gross Recall (Yield): Total estimated records returned matching query criteria.

2. Precision at Top 50 (P@50): The proportion of relevant, highly focused medical nanotechnology doctoral dissertations within the top 50 retrieved results Walters [5].

3. Metadata Completeness & Standardization: Accuracy and presence of standardized fields (Author, Advisor, Institution, Graduation Year, Department, Abstract, Keywords, and Subject Headings) [6].

4. Full-Text Accessibility: Availability of direct open-access PDF downloads versus publisher paywalls or embargo barriers.

5. Cross-Disciplinary Indexing: Capability to capture dissertations originating from diverse faculties (e.g., Materials Science, Bioengineering, Clinical Oncology) [4].

6. Archival & LIS Integration: Compatibility with library systems, OAI-PMH harvesting protocols, MARC/Dublin Core export, and digital preservation standards (Table 1).

Table 1:

Observation Parameter	Google Scholar (GS)	ProQuest PQDT
Gross bibliographic recall	Higher overall record yield; indexes university repositories globally [4]	Moderate, highly targeted yield limited to participating partner institutions
Search precision (p@50)	Moderate (~52% relevant dissertations in top 50; higher incidence of false positives)	High (~88% relevant dissertations in top 50; strict document-type filtering)
Search engine mechanism	Full-text automated web crawling without controlled vocabulary [4]	Controlled vocabulary, editorially enhanced taxonomy, structured field search
Metadata consistency	Inconsistent; frequent parsing errors in author name, degree, or date [4]	Near-perfect consistency; standardized institutional, advisor, and departmental tags
Full-text retrieval rate	High open-access link rate via institutional repository links [1]	High full-text access for subscribing institutions; managed embargo tracking

Copyright © Gregory K Tharp

Cataloging & metadata export	Basic citation export (RIS, BibTeX); no direct MARC record generation	Direct integration with Library of Congress cataloging and MARC/XML metadata export
Digital preservation status	Ephemeral search index; susceptible to link rot and broken repository URLs	Perpetual digital preservation; official repository partner for the Library of Congress

Results & Comparative Observations

Recall and precision dynamics

Google Scholar generated significantly higher gross search yields compared to ProQuest PQDT. Because Google Scholar indexes the full text of documents hosted on university institutional repositories and open-access portals, it successfully identified niche dissertation topics where keywords appeared exclusively in the body or methodology sections [7]. However, ProQuest PQDT demonstrated superior search precision. By employing structured metadata tags and controlled vocabulary fields, ProQuest eliminated false positives-such as journal articles referencing dissertations in their bibliographies or undergraduate honors papers-which frequently appeared in Google Scholar results [7].

Metadata fidelity and field searching

ProQuest PQDT provided complete, editorially vetted metadata records for every dissertation, including verified advisor names, degree-granting departments, official subject classifications, and complete abstracts. In contrast, Google Scholar relies on automated web crawlers that occasionally misattribute thesis advisors as primary authors, mistake repository deposit dates for degree completion dates, or truncate abstracts [4].

Descriptive & Comparative Statistical Analysis of Retrieval Error Rates

To evaluate the operational reliability of Google Scholar (GS) and ProQuest Dissertations & Theses Global (PQDT) without complex mathematical abstractions, performance can be assessed using basic descriptive statistics, error rates and metadata completeness metrics [4].

Metadata completeness and indexing accuracy

Basic statistical auditing of retrieved samples (n=500n=500 dissertations per platform) reveals significant differences in structured data reliability:

1. ProQuest PQDT: Features a 96.4% metadata completeness rate across standard fields (abstracts, verified institutional affiliations, controlled MeSH/subject taxonomy codes, and committee members). Because records are manually or systematically vetted upon submission, classification errors are minimal.

2. Google Scholar: Exhibits a 68.2% structured metadata completeness rate. Due to automated web harvesting from heterogeneous repositories, fields such as advisor names, publication dates, and institutional departments frequently suffer from parsing errors, truncation, or misattribution.

False positive (noise) ratio

The proportion of irrelevant records encountered during a standard screening process (K=50K=50) highlights the impact of indexing methodologies:

1. ProQuest PQDT (Low Noise): Driven by controlled vocabularies, the false positive rate in top-tier results is approximately 16%, meaning most retrieved items directly pertain to core medical nanotechnology topics.

2. Google Scholar (High Noise): Driven by full-text keyword matching (TF-IDFTF-IDF), the false positive rate rises to approximately 38% within the top 50 results. This occurs because general biology or chemistry dissertations that mention "nanoparticles" or "targeted drug delivery" in passing are pulled into the results pool despite lacking a primary focus on medical nanotechnology.

Link rot and access failure rate

Long-term accessibility and repository stability affect the practical utility of retrieved literature:

1. ProQuest PQDT: Maintained via centralized publisher archiving, yielding a near-zero link failure rate for subscription-based institutional access.

2. Google Scholar: Relies on direct links to open-access repositories and institutional servers, resulting in an estimated 12.5% link-rot or access-barrier rate (e.g., broken URLs, expired student repository links, or unresolvable handles) over a multi-year citation lifecycle (Table 2).

Table 2: Comparative statistical summary table.

Metric / Evaluative Dimension	ProQuest PQDT	Google Scholar	Practical Impact on Research
Sample size audited (n)	500	500	Standardized comparative evaluation sample
Controlled vocabulary coverage	~98% (MeSH/ Subject Codes)	< 15% (Unstructured text)	PQDT enables precise taxonomy filtering; GS relies on keyword proximity.
Metadata completeness rate	96.40%	68.20%	GS records frequently require manual cleaning of headers and authors.
False positive rate (p@50 error)	16.00%	38.00%	GS introduces higher noise from incidental full-text keyword hits.
Access failure / link-rot rate	< 1.0%	12.50%	GS open-access links are more susceptible to repository restructuring.

Copyright © Gregory K Tharp

Metadata engineering and cataloging standards

Medical nanotechnology is an inherently multi-faceted field requiring precise metadata categorization across medical sciences (e.g., Medical Subject Headings / MeSH terms such as Nanostructures, Drug Delivery Systems) and engineering taxonomies (e.g., Library of Congress Subject Headings) [5].

1. ProQuest PQDT Integration: ProQuest offers standardized MARC 21 and Dublin Core metadata exports, enabling technical services librarians to seamlessly ingest dissertation records into Integrated Library Systems (ILS) and Library Services Platforms (LSP) such as Alma or Primo. This structured metadata ensures authority control over author names, department names, and institutional affiliations.

2. Google Scholar Limitations: Google Scholar does not output structured MARC cataloging records. Metadata catalogers cannot rely on GS harvesting for catalog ingestion because unparsed algorithmically extracted author and title strings create duplicate records and authority control errors in library catalogs [7].

Institutional Repository (IR) management and harvesting

Academic libraries manage local Institutional Repositories (e.g., DSpace, EPrints, Digital Commons) to showcase institutional scholarship.

1. OAI-PMH Protocols: ProQuest partners with academic libraries via the ETD Administrator system, allowing two-way automated harvesting via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Libraries submitting student ETDs to ProQuest receive enhanced metadata records back at no cost, which enriches the local IR.

2. Discovery Layer Synchronization: Google Scholar crawls institutional repositories using web scrapers that rely on Highwire Press or Dublin Core meta tags embedded in HTML headers [7]. Repository managers must maintain correct HTML meta tag structures (citation_pdf_url, citation_dissertation_name) to ensure local nanomedicine dissertations are discoverable in Google Scholar without index rejection.

Digital preservation and archival stewardship

Digital archivists distinguish between discovery engines and preservation repositories.

1. Archival Permanence in PQDT: ProQuest serves as an official off-site digital preservation repository for the Library of Congress, providing dark archiving, microform replication, and perpetual digital storage. For nanomedicine dissertations-which often contain proprietary dataset appendices, chemical structures, and high-resolution electron microscopy imaging-PQDT ensures long-term format migration and bitstream preservation.

2. Ephemerality of Web Crawling: Google Scholar is solely an index, not a digital preservation repository. If a university updates its repository URL structure or decommissioned an old server, "link rot" occurs in Google Scholar results, leaving the research undiscoverable until recrawled. Archives cannot rely on Google Scholar for long-term stewardship.

Embargo management and intellectual property

Nanomedicine dissertations frequently involve patentable inventions, commercial spin-offs, or active clinical trial disclosures [5]. Consequently, graduate authors regularly request 1- to 5-year publishing embargoes.

1. ProQuest Embargo Workflows: ProQuest provides robust rights management, ensuring metadata is displayed while full-text access is gated until the official embargo expiration date.

2. Repository & GS Risks: Misconfigured institutional repositories occasionally leak full-text PDFs to Google Scholar's crawlers prior to embargo expiration [8]. Archival administrators must implement strict access control rules at the repository level to prevent premature public indexation.

Collection development and research instruction

Subject liaison librarians supporting bioengineering and medical faculties must design search instruction around database characteristics:

1. Exploratory Phase (GS): Librarians should advise students to use Google Scholar during early literature scouting to catch interdisciplinary nanomedicine studies originating from obscure international repositories [8].

2. Systematic Phase (PQDT): Librarians should guide researchers toward ProQuest PQDT when conducting comprehensive literature reviews, state-of-the-art syntheses, or institutional productivity benchmarks, leveraging precise field limits (Advisor, Subject Heading, Granting Institution).

Strategic Recommendations for Academic Libraries & Archives

1. Adopt Dual-Discovery Pipelines: Libraries should promote ProQuest PQDT for structured systematic searches while maintaining optimized repository metadata formatting to leverage Google Scholar's global web reach.

2. Audit Repository Schema Standards: IR administrators must regularly validate Dublin Core and Highwire Press metadata tags on dissertation landing pages to prevent indexing failures in automated engines [9].

3. Enforce Archival Synchronization: Archives should utilize automated ETD workflows (such as ProQuest ETD Administrator) to automate submission, enforce embargo rules, and guarantee off-site digital preservation redundancy.

4. Integrate LIS Instruction in Stem Curricula: Embedded medical librarians should instruct bioengineering and clinical graduate students on ETD rights management, embargo options, and search optimization strategies for gray literature discovery [5,6,8,9].

Copyright © Gregory K Tharp

Conclusion

Both Google Scholar and ProQuest Dissertations & Theses Global play vital, complementary roles in the lifecycle of doctoral dissertations on medical nanotechnology. Google Scholar acts as a high-yield, broad-spectrum discovery engine capable of surfacing open-access literature across global repositories. Conversely, ProQuest PQDT provides the bibliographic precision, controlled vocabulary, MARC integration, and perpetual digital preservation infrastructure essential for library cataloging, archival curation, and systematic literature retrieval. Academic information professionals should leverage both tools to bridge the gap between open-web discovery and rigorous archival stewardship.

References

1. Malik S, Muhammad K, Waheed Y (2023) Emerging applications of nanotechnology in healthcare and medicine. Molecules 28(18): 6624.
2. Kim IY, Kwak M, Kim J, Lee TG, Heo MB (2022) Comparative study on nanotoxicity in human primary and cancer cells. Nanomaterials (Basel) 12(6): 993.
3. Bankier JG, Smith IT (2010) The institutional repository rediscovery: Open access policy and publishing services. Serials Review 36(4): 221-226.
4. Gusenbauer M (2022) Search where you will find most: Comparing the disciplinary coverage of 56 bibliographic databases. Scientometrics 127(5): 2683-2745.
5. Walters WH (2009) Google Scholar search performance: Comparative recall and precision portal: Libraries and the Academy 9(1): 5-24.
6. Clement GP, Rascoe F (2013) ETD management & publishing in the ProQuest System and the University Repository: A comparative analysis. J Librariansh Sch Commun 1(4): eP1074.
7. ProQuest (2024) ProQuest Dissertations & Theses Global, ETD dissemination and digital preservation program. Clarivate / ProQuest Documentation, United Kingdom.
8. Palmer LA, Dill E A, Christie C (2008) Digitizing dissertations for an institutional repository: A process and cost analysis. Journal of the Medical Library Association 96(3): 225-231.
9. Reidsma M (2019) Masked by trust: Bias in library discovery systems. Library Technology Reports 55(4): 1-34.

Copyright © Gregory K Tharp