

Preserving a Fading Archive: Public Perceptions and the Case for a National Digital Genealogical Repository in India

Dr. (Prof.) Mandeep Kaur Kochar¹, Ishwar Singh², Dr. Sahibpreet Singh Kochar³,
Rahul Pawar⁴

¹College of Education, Bombay Teachers Training College, Mumbai, India

²Department of Political Science and Mechanics, The Kalgidhar Society, Baru Sahib, India

³Department of Emerging Technologies, The Kalgidhar Society, Baru Sahib, India

⁴Department of Computer Science and Engineering, Amrita School of Computing, Coimbatore Campus, Amrita Vishwa Vidyapeetham University, Coimbatore - 641 112, Tamil Nadu, India

Abstract

India's pilgrimage centres contain hereditary genealogical registers that record family relationships, places of origin, ritual visits, and other historical information. These registers are culturally important but remain materially vulnerable because they are maintained in dispersed, locally controlled collections. This study examines public perceptions of digitizing such records and evaluates support for a proposed National Digital Genealogical Repository (NDGR). The research uses a structured cross-sectional survey design with eight demographic variables, thirty-two five-point Likert statements organized into eight conceptual constructs, and three open-ended questions. For manuscript development, the numerical dataset and statistical outputs presented here are rather than observations collected from human participants; consequently, the findings should be treated as an analytical prototype and not as evidence of actual population attitudes. The analysis indicates high perceived heritage value ($M = 3.68$) and positive support for an NDGR ($M = 3.49$), alongside comparatively low awareness ($M = 2.81$). Awareness is positively associated with support ($r = .43$, $R^2 = .185$, $p < .001$), while privacy and trust show a weaker relationship with support ($r = .28$). The proposed governance model therefore emphasizes awareness, informed consent, security, community participation, recognition of hereditary genealogists, and strict limits on administrative use. The paper concludes that a future empirical study should replace the dataset with ethically collected primary observations before any substantive claims about public opinion are published.

Keywords: digital preservation; genealogy; cultural heritage; archival digitization; India; pilgrimage records; data governance

1. Introduction

When families visit major Hindu pilgrimage centres for ancestral rites, they may encounter hereditary genealogists who maintain registers containing family names, places of origin,

*Corresponding Author Email: skochar644@gmail.com

Published: 27 August 2026

DOI: <https://doi.org/10.70558/IJSSR.2026.v3.i4.301265>

Copyright © 2026 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

relationships, and records of previous visits. Haridwar is particularly well documented as a pilgrimage centre in which pandas have historically maintained genealogical records, while similar traditions occur at other pilgrimage locations. James G. Lochtefeld's fieldwork on Haridwar describes the pandas as a distinctive occupational community within the pilgrimage economy and provides important context for understanding the relationship between genealogists, pilgrims, and the historical development of the site (Lochtefeld, 2010). The FamilySearch collection titled India, Hindu Pilgrimage Records, 1194–2015 also demonstrates that substantial collections of pilgrimage records exist and that their organization is closely connected to individual pandits and geographical or family groupings (FamilySearch, 2025).

The preservation problem is not simply the survival of paper. A genealogical register can be simultaneously a documentary object, a family resource, a source for historical research, and a record containing information about identifiable people. Digitization can reduce the risk associated with dependence on a single physical copy, but scanning alone does not constitute digital preservation. The International Federation of Library Associations and Institutions (IFLA) and the International Council on Archives (ICA) emphasize planning, technical standards, quality control, metadata, and long-term management as integral elements of digitization programmes (McIlwaine et al., 2002). The Open Archival Information System reference model likewise treats preservation as an ongoing system of information management rather than a one-time conversion activity (Consultative Committee for Space Data Systems [CCSDS], 2024).

The proposed National Digital Genealogical Repository (NDGR) is therefore a governance problem as much as a technical one. A repository could increase the discoverability and survival of records, but it could also alter relationships between hereditary custodians, families, researchers, and public institutions. The records may contain personal and relational information, and any digital system must therefore distinguish legitimate preservation and research purposes from inappropriate uses. India's Digital Personal Data Protection Act, 2023 provides an important contemporary legal context because it establishes a statutory framework concerning the processing of digital personal data, including notice, consent, obligations of data fiduciaries, and rights of data principals (Government of India, 2023).

This paper asks how a survey-based research design could measure public awareness, heritage value, support for an NDGR, perceived benefits, willingness to participate, privacy and trust, recognition of genealogists, and attitudes toward administrative limitations. Four hypotheses are specified to connect these constructs. The manuscript also demonstrates how descriptive statistics, internal-consistency reliability, correlations, regression, and group comparison can be used to analyze the proposed instrument. The important methodological qualification is that the numerical results in this version are. They are retained from the supplied manuscript as an illustrative dataset, not represented as observations genuinely collected from 350 respondents. This distinction is essential for research integrity.

The remainder of the paper reviews the documentary and governance context, states the research objectives and hypotheses, describes the survey instrument and analytical procedures, presents the results, discusses their implications for repository design, identifies limitations, and proposes a framework for a subsequent study based on genuine primary data.

2. Background: India's Traditional Genealogical Record-Keeping

Traditional genealogical registers associated with pilgrimage centres are distributed across multiple locations and custodial relationships. Haridwar provides the clearest documented example in the literature. Lochtefeld (2010) describes the *pandas* of Haridwar and the changing social and economic conditions in which pilgrimage and record-keeping developed. The records are not best understood as a single centrally administered archive; rather, they are dispersed among hereditary or locally established custodians whose knowledge of family origins and relationships has historically helped pilgrims locate the relevant register.

The documentary significance of these records extends beyond the immediate purpose for which individual entries were created. Genealogical information can provide evidence for family history and can contribute to broader historical investigations of migration, residence, kinship, and social change. At the same time, the information may concern living people and their relatives. The dual status of genealogical registers as heritage resources and repositories of personal information means that preservation policy must address both cultural value and data governance.

The physical condition of manuscript heritage also makes digitization relevant. UNESCO's Charter on the Preservation of Digital Heritage emphasizes that documentary materials converted into digital form require purposeful maintenance, management, and preservation if they are to remain accessible to future generations (UNESCO, 2003). IFLA's digitization guidance similarly emphasizes technical specifications, metadata, quality control, rights considerations, and project planning rather than treating digitization as a simple scanning exercise (McIlwaine et al., 2002). These principles are applicable to a future NDGR, although the repository would need to be adapted to the dispersed custodianship and cultural sensitivities of Indian genealogical records.

3. Literature Review

3.1 Digital Preservation Of Archival And Manuscript Heritage

Digital preservation is an active lifecycle process involving the capture, description, storage, integrity management, access, and future migration of digital objects. The OAIS reference model provides a widely used conceptual framework for understanding trustworthy digital preservation, including the responsibilities associated with ingest, archival storage, data management, preservation planning, and access (CCSDS, 2024). For manuscript heritage, IFLA and ICA recommend that digitization projects be planned around institutional objectives, technical quality, metadata, staffing, sustainability, and access conditions (McIlwaine et al., 2002). UNESCO likewise identifies long-term continuity, authenticity, legal frameworks, and stakeholder cooperation as essential to preservation of documentary heritage in digital form (UNESCO, 2003).

These principles have two implications for an NDGR. First, digital images should not be treated as self-describing objects. They require metadata that explains provenance, relationships, dates, formats, rights, and preservation information. Second, a repository must be designed for

continuity. Hardware, software, file formats, storage environments, and organizational responsibilities change over time, so preservation cannot end when a scanning project finishes.

3.2 Genealogical Records As Heritage And Personal Data

Genealogical records have a distinctive documentary character because they connect individuals across generations. FamilySearch's catalogued collection of Indian Hindu pilgrimage records shows that these records can include names, family relationships, native places, dates of visits, ceremonies, and other contextual information (FamilySearch, 2025). Lochtefeld's study of Haridwar provides a scholarly account of the pandas and the social organization surrounding the pilgrimage centre (Lochtefeld, 2010). Together, these sources demonstrate why genealogical records can be valuable for both family history and historical research.

The same characteristics create privacy and governance risks. A digitized record can be copied, searched, indexed, linked to other datasets, or accessed remotely at a scale that differs substantially from consultation of a physical register held by a local custodian. The Digital Personal Data Protection Act, 2023 is therefore relevant to the design of any future repository that processes digital personal data in India (Government of India, 2023). A repository should establish clear purposes, access rules, consent mechanisms where appropriate, security controls, retention rules, and procedures for responding to data-subject rights and other legal requirements.

3.3 Participation, Trust, And Community Governance

Community participation is particularly important when records are held by people whose authority derives from customary practice rather than a centralized archival institution. Research on community archives shows that trust-building, community ownership, participatory description, and collaborative practice can challenge conventional custodial models (Stevens et al., 2018). Jansson (2020) similarly examines participatory archival arrangements in which professional authority is balanced with participant influence. These studies suggest that an NDGR should not be designed as a unilateral state takeover of privately or communally held records.

The governance implication is that hereditary genealogists should be treated as stakeholders in preservation rather than merely as sources from which materials are extracted. Their expertise can contribute to identification, contextual description, interpretation, access decisions, and validation. A repository could therefore adopt a federated or post-custodial model in which high-quality digital preservation infrastructure is centrally coordinated while custodians retain defined roles and rights over source materials and access conditions.

3.4 Research Gap

The supplied manuscript identifies a gap between descriptive discussion of Indian genealogical registers and quantitative measurement of public attitudes toward their digitization. That gap is plausible as a research question, but the present dataset cannot establish the empirical claim that no previous quantitative study exists. A future systematic literature review would be required to substantiate a priority or originality claim. Accordingly, this revised manuscript

does not describe the study as the first quantitative investigation. Instead, it presents the survey instrument and analysis as a research prototype that can be tested with genuine primary data.

4. Research Objectives And Hypotheses

The study has four objectives. The first is to measure public awareness of traditional genealogical record-keeping and the cultural value attached to it. The second is to assess support for a National Digital Genealogical Repository and the benefits associated with such a repository. The third is to examine privacy, institutional trust, recognition of genealogists, and perceived limits on administrative use. The fourth is to translate the measured relationships among these constructs into governance requirements for a future repository.

H1 states that awareness of the genealogical tradition is positively associated with support for an NDGR. H2 states that perceived cultural-heritage value is positively associated with support for digitization. H3 states that privacy and trust concerns are only weakly associated with support for digitization, indicating that concern about data handling may coexist with support for preservation. H4 states that willingness to participate does not differ significantly between diaspora (NRI) and resident respondents.

5. Methodology

5.1 Research Design

A quantitative, cross-sectional survey design was specified because the research questions concern the distribution and interrelationship of attitudes across respondents. The instrument is structured to support descriptive analysis, internal-consistency reliability assessment, Pearson correlations, an independent-samples t-test, and simple linear regression. The design is appropriate for a future primary-data study provided that sampling, recruitment, consent, data protection, and ethical review are implemented before data collection.

Important data-status disclosure: the numerical observations represented in this manuscript are. The supplied manuscript itself labels the tables as illustrative simulated data. Therefore, this version does not claim that 350 human participants actually completed the questionnaire. The results are retained solely to demonstrate how the proposed study could be reported. They must be replaced by genuine, ethically collected primary data before the manuscript is submitted as an empirical research article.

5.2 Instrument

The proposed questionnaire contains three parts. Part A collects eight demographic and background variables: age group, gender, highest education, residential area, state or union territory, residence outside India, previous visitation to a pilgrimage centre for ancestral rites, and whether the respondent's family has records with a hereditary genealogist. Part B contains thirty-two five-point Likert statements, with 1 representing strongly disagree and 5 representing strongly agree. The statements are organized into eight conceptual constructs: awareness, heritage value, support for the NDGR, perceived benefits, willingness to participate, privacy and trust, recognition of genealogists, and use and limitations. Part C contains three open-ended questions concerning perceived benefits, concerns, and suggestions.

Two privacy-related statements are negatively worded and would be reverse-coded when a composite score is calculated so that a higher score consistently indicates a more trusting orientation. The present item table reports the means for the original wording, while the construct-level privacy-and-trust score is based on the recoded items. Before genuine data collection, the instrument should undergo content validation, cognitive testing, pilot administration, and, if sample size permits, factor-analytic assessment. High Cronbach's alpha alone does not establish construct validity; methodological literature cautions that alpha should be interpreted alongside dimensionality and item behavior (Tavakol & Dennick, 2011).

5.3 Sampling And Data Collection

The proposed empirical study targets adults aged 18 years or above and uses an anonymous self-administered questionnaire. A sample of 350 was specified in the supplied manuscript. For a genuine study, the sampling frame and recruitment channels must be documented, including how resident and diaspora participants are reached and how duplicate or ineligible responses are excluded. Informed consent should be obtained before participation, and the research protocol should receive the institutional ethics approval required by the host institution and applicable rules.

Because the current dataset is, no claim is made here that the described questionnaire was actually administered to 350 people. The demographic profile and all inferential results below are illustrative values retained for analytical demonstration.

5.4 Data Analysis

Categorical variables are summarized using frequencies and percentages, while Likert constructs are summarized using means and standard deviations. Internal consistency is assessed using Cronbach's alpha, with 0.70 treated as a commonly used reference point rather than an absolute validity criterion (Tavakol & Dennick, 2011). Pearson correlations are used to examine associations among constructs. A simple linear regression examines the relationship between awareness and support for the NDGR. An independent-samples t-test compares willingness to participate between NRI and resident respondents. Statistical significance is evaluated at $p < .05$. Because the numerical dataset, these p-values and effect estimates are illustrative rather than inferential evidence about an actual population.

6. Results

Table 1. Respondent Profile in The Dataset (n = 350)

Characteristic	Category	n (%)
Age	18–25	115 (32.9%)
	26–35	89 (25.4%)
	36–50	76 (21.7%)
	51–65	46 (13.1%)

	Above 65	24 (6.9%)
Gender	Male	191 (54.6%)
	Female	151 (43.1%)
	Prefer not to say	8 (2.3%)
Education	Higher secondary	45 (12.9%)
	Undergraduate	153 (43.7%)
	Postgraduate	102 (29.1%)
	Doctorate/Research	30 (8.6%)
	Other	20 (5.7%)
Residential area	Urban	199 (56.9%)
	Semi-urban	94 (26.9%)
	Rural	57 (16.3%)
Living outside India	Yes	55 (15.7%)
	No	295 (84.3%)
Visited a centre for ancestral rites	Yes	222 (63.4%)
	No	81 (23.1%)
	Not sure	47 (13.4%)
Family has records with a genealogist	Yes	111 (31.7%)
	No	97 (27.7%)
	Not sure	142 (40.6%)

The profile is concentrated in younger, urban, and formally educated categories. A majority of the illustrative respondents report having visited a pilgrimage centre for ancestral rites, while a plurality is uncertain whether their family has records with a genealogist. These values are descriptive properties of the dataset and should not be interpreted as estimates of the Indian population.

Table 2. Construct Means, Standard Deviations, And Internal Consistency

Construct	Mean	SD	α
Awareness	2.81	0.67	0.90

Heritage value	3.68	0.65	0.89
Support for NDGR	3.49	0.71	0.91
Perceived benefits	3.42	0.67	0.90
Willingness to participate	3.23	0.69	0.91
Privacy & trust (recoded)	3.07	0.65	0.89
Recognition of genealogists	3.49	0.61	0.89
Use & limitations	3.55	0.66	0.90

In the dataset, heritage value has the highest mean ($M = 3.68$, $SD = 0.65$), whereas awareness has the lowest ($M = 2.81$, $SD = 0.67$). Support for the NDGR is positive relative to the five-point scale midpoint ($M = 3.49$, $SD = 0.71$). The privacy-and-trust score is close to the midpoint ($M = 3.07$, $SD = 0.65$), suggesting illustrative ambivalence rather than either strong confidence or strong distrust. All eight alpha coefficients exceed .89. These coefficients demonstrate internal consistency within the item structure, but they do not establish that the constructs are valid measures in a real population.

Table 3. Item Means and Standard Deviations in The Dataset

Item	Mean	SD
B1 Familiar with genealogist tradition	2.84	0.77
B2 Understand recorded information	2.78	0.76
B3 Aware records span generations	2.81	0.77
B4 Know how to locate family records	2.79	0.75
C1 Important part of cultural heritage	3.67	0.74
C2 Historical value beyond families	3.69	0.74
C3 Loss would be civilizational loss	3.66	0.75
C4 Important to preserve for future	3.70	0.77
D1 Should be digitized and preserved	3.48	0.80
D2 NDGR would be valuable	3.49	0.78
D3 Digitization protects better than manuscripts	3.50	0.84
D4 Government should invest	3.50	0.77
E1 Easier to trace family history	3.44	0.77
E2 Supports academic research	3.43	0.77

E3 Encourages heritage tourism	3.41	0.76
E4 Helps diaspora reconnect	3.38	0.75
F1 Interested in accessing records	3.24	0.78
F2 Willing to contribute information	3.23	0.78
F3 Willing to upload documents	3.22	0.76
F4 Would recommend to others	3.24	0.77
G1 Concerned about privacy (reverse-coded)	3.04	0.77
G2 Worry about data misuse (reverse-coded)	3.06	0.75
G3 Access should require consent	3.11	0.75
G4 Trust institutions to manage securely	3.07	0.76
H1 Genealogists should be compensated	3.49	0.71
H2 Recognized as heritage custodians	3.48	0.70
H3 Collaboration not takeover	3.50	0.71
H4 Training and professional recognition	3.50	0.71
I1 Supporting evidence for kinship	3.55	0.77
I2 Not primary proof of citizenship	3.54	0.74
I3 No disadvantage if records absent	3.57	0.76
I4 Participation must be voluntary	3.56	0.75

The item-level pattern is consistent with the construct-level results. Awareness items cluster below the midpoint, whereas heritage-value items are comparatively high. Support items are slightly above the midpoint, and the use-and-limitations items are also relatively high. The responses therefore represent a public that, under the assumptions built into the illustrative dataset, places value on preservation while simultaneously endorsing voluntary participation, consent, and restrictions on administrative use.

Table 4. Pearson Correlations Among Constructs in the Dataset

Construct	Heritage	Support	Benefits	Willingness	Privacy/Trust	Recognition	Use Limits	Awareness
Heritage	1.00	0.45	0.41	0.43	0.16	0.46	0.41	0.40
Support	0.45	1.00	0.55	0.52	0.28	0.50	0.36	0.43
Benefits	0.41	0.55	1.00	0.44	0.19	0.46	0.35	0.43

Willingness	0.43	0.52	0.44	1.00	0.23	0.42	0.38	0.41
Privacy/Trust	0.16	0.28	0.19	0.23	1.00	0.25	0.17	0.17
Recognition	0.46	0.50	0.46	0.42	0.25	1.00	0.36	0.37
Use Limits	0.41	0.36	0.35	0.38	0.17	0.36	1.00	0.32
Awareness	0.40	0.43	0.43	0.41	0.17	0.37	0.32	1.00

Support for the NDGR shows moderate positive associations with perceived benefits ($r = .55$), willingness to participate ($r = .52$), recognition of genealogists ($r = .50$), heritage value ($r = .45$), and awareness ($r = .43$). The privacy-and-trust construct has weaker correlations with the other constructs, with the largest association being $r = .28$ with support. This pattern is consistent with the hypothesis that positive attitudes toward digitization can coexist with reservations about data governance. However, correlation alone cannot establish that privacy and trust constitute a statistically distinct latent factor; factor analysis would be needed to test that claim.

Table 5. Hypothesis Tests Based on The Dataset

Hypothesis	Statement	result
H1	Awareness is positively associated with support for the NDGR.	Supported: $\beta = .43$, $r = .43$, $R^2 = .185$, $p < .001$
H2	Heritage value is positively associated with support.	Supported: $r = .45$, $p < .001$
H3	Privacy/trust is only weakly associated with support.	Supported: $r = .28$, $p < .001$
H4	Willingness does not differ significantly between NRI and resident respondents.	Supported as a null result: NRI $M = 3.14$; resident $M = 3.25$; $t = -1.15$; $p = .25$

The simple regression of support on awareness is statistically significant in the dataset, with awareness accounting for approximately 18.5% of the variance in support ($R^2 = .185$). Because the analysis is based on a dataset, the coefficient should be interpreted only as an internally consistent demonstration of the proposed model. The positive association between heritage value and support is also consistent with H2. The privacy-and-trust relationship is weaker than the principal positive associations, supporting H3 as operationalized here. Finally, the illustrative t-test does not show a statistically significant difference in willingness between NRI and resident groups, supporting H4 as a null result.

7. Discussion

The results suggest a potentially important relationship between cultural value and public awareness. Heritage value is the strongest construct in the illustrative dataset, while awareness is the weakest. At the same time, awareness is positively associated with support for an NDGR. If this pattern were reproduced in genuine primary data, it would imply that preservation policy should not assume that support automatically follows from the existence of heritage value. Public communication would need to explain what the records are, where they are held, how they can be consulted, and why their preservation matters.

The second implication concerns privacy and trust. The privacy-and-trust construct is only weakly related to support, suggesting that respondents could support digitization while simultaneously demanding safeguards. This is consistent with the governance logic of documentary heritage preservation: accessibility and protection of sensitive information must be balanced rather than treated as mutually exclusive objectives. UNESCO explicitly recognizes the importance of protecting sensitive and personal information while expanding access to digital heritage (UNESCO, 2003). India's data-protection framework provides an additional legal context for designing notices, consent processes, security measures, and data-subject rights into any future system (Government of India, 2023).

A third implication concerns the role of hereditary genealogists. The respondents express relatively high support for compensation, recognition, training, and collaboration rather than institutional takeover. If confirmed through genuine research, this would favor a participatory governance model. Community-archive research emphasizes that trust and engagement can be strengthened when communities retain meaningful ownership and influence over archival practices (Stevens et al., 2018). Jansson (2020) similarly shows the value of controlled power-sharing between professional archivists and participants. For an NDGR, hereditary genealogists could therefore have formal representation in governance, metadata standards, access policies, digitization protocols, and dispute-resolution mechanisms.

The fourth implication is the importance of purpose limitation. The responses favor the use of records as supporting evidence for family relationships while rejecting their use as primary proof of citizenship and emphasizing voluntary participation. This is a significant governance distinction. A heritage repository designed to preserve family records should not quietly become an identity-verification database. Such function expansion could alter the risk profile of the system, undermine trust, and create incentives for coercive participation. A future NDGR should therefore have an explicit purpose statement, documented access categories, audit controls, and a prohibition on secondary administrative uses that have not been separately authorized and ethically assessed.

8. Implications And Recommendations

An awareness-led repository should make public education a core programme rather than a promotional activity added after technical development. Outreach should explain the historical role of genealogists, the kinds of information contained in registers, the difference between preservation copies and public access copies, and the choices families have over access. This

approach would also allow future research to measure whether awareness interventions actually change knowledge and willingness to participate.

Consent and security should be embedded in the repository architecture. A future system should distinguish between digitization consent, preservation custody, research access, family access, and public discovery. Sensitive fields may require restricted access even when a digital image is preserved. Security controls should include authentication, authorization, encryption where appropriate, audit logging, backup, integrity checking, and documented incident-response procedures. These requirements are consistent with the broader preservation principles of OAIS and UNESCO and should be adapted to India's legal environment (CCSDS, 2024; Government of India, 2023; UNESCO, 2003).

Hereditary genealogists should be recognized as heritage custodians and knowledge partners. Their participation should include fair compensation for digitization-related work, training in handling and imaging, participation in metadata development, and representation in governance. The repository should avoid a model in which a central institution merely extracts copies from local custodians and then excludes them from decision-making.

The NDGR should also be purpose-limited. Its policy framework should define the repository primarily as a cultural-heritage and research resource. Any use for legal, administrative, identity, or citizenship purposes should be governed separately and should not be inferred merely from the existence of a digital record. The principle that absence of a record must not create disadvantage is especially important because the survival and completeness of genealogical registers are uneven.

Finally, implementation should be phased. A pilot involving one or two participating centres could test imaging quality, metadata fields, multilingual description, access controls, consent procedures, data security, and community governance before national expansion. IFLA's digitization guidance supports project planning and quality control, while OAIS provides a framework for thinking about long-term preservation responsibilities (McIlwaine et al., 2002; CCSDS, 2024). A pilot would also provide an opportunity to compare stated willingness with actual participation.

9. Limitations And Future Research

The most important limitation is the data status. The numerical findings in this manuscript are and cannot support claims about the attitudes of Indian adults. They should not be cited as evidence of public opinion, and the manuscript should not be submitted as an empirical survey study until genuine primary data have been collected and analyzed. The second limitation is that the proposed survey is cross-sectional and self-reported; even with genuine data, it would establish associations rather than causal relationships. The third limitation concerns sampling. An online sample could under-represent rural populations, older adults, people with limited digital access, and hereditary genealogists themselves.

Future research should begin with a genuine pilot survey using a documented sampling strategy and an ethics-approved protocol. The instrument should be content-validated by scholars of archival studies, genealogy, data governance, and Indian heritage, and reviewed by hereditary

genealogists. A pilot study should test item comprehension and response distributions before full administration. Exploratory or confirmatory factor analysis should be considered to assess whether the proposed eight-construct structure is empirically supported. Qualitative interviews with genealogists, archivists, families, legal experts, and policymakers should complement the survey because many repository-governance questions cannot be reduced to Likert-scale attitudes.

A subsequent study should also compare different access models, such as family-only access, controlled researcher access, and public discovery metadata. Experimental or vignette-based designs could test whether explicit consent guarantees, genealogist representation, security information, or restrictions on administrative use change willingness to participate. Finally, a functioning pilot repository should be evaluated using actual participation, retrieval, metadata quality, preservation integrity, user satisfaction, and governance outcomes.

10. Conclusion

Traditional genealogical registers associated with Indian pilgrimage centres represent an important documentary resource whose preservation involves cultural, technical, legal, and community dimensions. Digitization can improve resilience and access, but it does not by itself resolve questions of custodianship, privacy, consent, authenticity, or long-term preservation. The proposed NDGR therefore requires a governance architecture that treats hereditary genealogists and families as participants rather than merely as sources of data.

The analysis presented in this manuscript demonstrates a feasible quantitative framework for studying these issues. In the illustrative dataset, heritage value is relatively high, awareness is comparatively low, and support for an NDGR is positively associated with awareness and heritage value. Privacy and trust show a weaker relationship with support, while willingness to participate does not differ significantly between the illustrative NRI and resident groups. These numerical patterns are useful for demonstrating the proposed analysis but are not evidence of actual public attitudes.

The central research requirement is therefore clear: the next stage must replace the dataset with ethically collected primary data. Only then can the hypotheses be evaluated as empirical claims and the proposed governance model be tested against the experiences and preferences of real participants. Until that work is completed, this manuscript should be regarded as a transparent research prototype rather than as evidence of completed primary research.

AI Assistance Disclosure

Generative artificial-intelligence tools were used to assist with manuscript restructuring, language editing, statistical-consistency checking, formatting, and reference verification during preparation of this document. The supplied manuscript and its stated research concept formed the basis of the work. The numerical results remain explicitly identified as, and no claim is made that the observations were collected from human participants. The authors independently verify every factual statement, source, citation, methodological detail, and result before submission and should ensure that the final manuscript complies with the journal's requirements concerning authorship and AI assistance.

References

- Consultative Committee for Space Data Systems. (2024). Reference model for an open archival information system (OAIS) (CCSDS 650.0-M-3). CCSDS. <https://ccsds.org/publications/allpubs/entry/3054/>
- FamilySearch. (2025). India, Hindu pilgrimage records, 1194–2015. FamilySearch. <https://www.familysearch.org/en/search/collection/1867930>
- Government of India. (2023). The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023). India Code. <https://www.indiacode.nic.in/handle/123456789/22037>
- Jansson, I.-M. (2020). Creating value of the past through negotiations in the present: Balancing professional authority with influence of participants. *Archival Science*, 20, 327–345. <https://doi.org/10.1007/s10502-020-09339-8>
- Lochtefeld, J. G. (2010). *God's gateway: Identity and meaning in a Hindu pilgrimage place*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195386141.001.0001>
- McIlwaine, J., Comment, J.-M., de Wolf, C., Peters, D., Justrell, B., Varlamoff, M.-T., & Koopman, S. (2002). Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives. *International Federation of Library Associations and Institutions*. <https://repository.ifla.org/items/5160f4c0-739d-44ad-a538-5ab466a962e6>
- Stevens, M., Flinn, A., & Shepherd, E. (2018). ‘A process where we’re all at the table’: Community archives challenging dominant modes of archival practice. *Archives and Records*, 39(1), 1–22. <https://doi.org/10.1080/01576895.2017.1377088>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- UNESCO. (2003). Charter on the preservation of digital heritage. UNESCO. <https://www.unesco.org/en/legal-affairs/charter-preservation-digital-heritage>

Appendix A. Survey Instrument

Part A collects age group, gender, highest education, residential area, state or union territory, residence outside India, previous family visitation to a pilgrimage centre for ancestral rites, and whether the family has records with a hereditary genealogist.

Part B uses a five-point scale from 1 = strongly disagree to 5 = strongly agree. Awareness items assess familiarity with the genealogist tradition, understanding of recorded information, awareness that records can span generations, and knowledge of how family records may be located. Heritage-value items assess cultural importance, historical value beyond individual families, the significance of loss, and preservation for future generations.

Support-for-NDGR items assess whether records should be digitized and preserved, whether an NDGR would be valuable, whether digitization can provide stronger preservation protection than vulnerable manuscripts, and whether public investment is appropriate. Perceived-benefit

items assess easier family-history research, academic research, heritage tourism, and diaspora reconnection.

Willingness-to-participate items assess interest in accessing records, willingness to contribute or update information, willingness to upload documents or photographs, and willingness to recommend the initiative. Privacy-and-trust items assess concern about privacy, concern about misuse, the need for family consent for access, and trust in institutions to manage records securely. The negatively worded concern items are reverse-coded for composite scoring.

Recognition-of-genealogists items assess financial compensation, recognition as heritage custodians, collaboration rather than institutional takeover, and training or professional recognition. Use-and-limitations items assess the value of records as supporting evidence for family relationships, rejection of the records as primary proof of citizenship, protection against disadvantage when records are absent, and voluntary participation.

Part C contains open-ended questions asking respondents to identify the main benefits of digitization, their principal concerns, and additional recommendations for the proposed repository.