International Journal of Classified Research Techniques & Advances (IJCRTA)
Multidisciplinary Peer-Reviewed Online Open Access Journal Impact Factor: 5.8
Volume 5, Issue 4, Oct-Dec 2025 ID: IJCRTA000217 ISSN: 2583-1801
Email: ijcrta@gmail.com Website: www.ijcrta.org

"Secure Archiving and Open Data Sharing: Methods and Risks"

Dr. Bhale Y.P

Author Affiliation:
Librarian, Arts & Science College, Chincholi (Limbaji), Tq. Kannad, Chh.Sambhajinagar.
Email.Yogeshbhale75@gmail.com

Citation of Article: Bhale, Y. P. (2025). Secure Archiving and Open Data Sharing: Methods and Risks. International Journal of Classified Research Techniques & Advances (IJCRTA) ISSN: 2583-1801, 5 (4), pg. 21-28. ijcrta.org

DOI: 10.5281/zenodo.17941545

Abstract:

As the global research landscape shifts toward open science, secure archiving and open data sharing have become central to responsible and impactful research practices. While open data accelerates discovery, increases reproducibility, and fosters interdisciplinary innovation, it also introduces critical risks such as security breaches, privacy violations, long-term preservation challenges, and ethical dilemmas. This extended article examines the principles, infrastructure, and techniques behind secure data archiving; explores global frameworks and policies promoting open data; discusses ethical, legal, and technological risks; and proposes comprehensive strategies for balancing openness with security. Through a review of contemporary literature and international standards, the article highlights the need for robust data governance, sustainable repositories, and advanced privacy-protection methods to ensure that research ecosystems remain both secure and transparent.

Keywords: Secure Data Archiving, Open Data Sharing, FAIR Principles, Digital Preservation

Introduction:

The digital revolution has transformed how scientific research is conducted, stored, and shared. Data generated across disciplines—from genomics and climate research to social sciences and engineering—has grown in volume, complexity, and strategic value. As a result, secure data archiving and open data sharing have become essential components of modern research infrastructure. International organizations, including the OECD (2015), UNESCO (2021), and the European Commission (2016) have called for greater openness to strengthen transparency, accelerate innovation, and ensure equitable access to publicly funded research. The FAIR principles (Findable, Accessible, Interoperable, Reusable), introduced by Wilkinson et al. (2016), further encourage institutions to design data systems that maximize reusability.

However, the shift toward openness is not without challenges. Research data may contain sensitive personal information, intellectual property, or cultural knowledge that must be protected. Moreover, insecure archiving systems expose institutions to data loss, cyber-attacks, and corruption. The tension between openness and security requires a nuanced, well-governed approach.

This article provides an in-depth examination of the methods used to achieve secure archiving, the frameworks that facilitate open data sharing, and the risks inherent in both processes. It concludes with best practices nd policy recommendations for harmonizing security with accessibility.

Review of Literature:

1. Data Sharing and Research Transparency:

Several studies highlight that secure archiving and open data sharing improve research transparency, reproducibility, and scientific integrity. Piwowar and Vision (2013) found that studies that publicly shared their datasets received significantly more citations, demonstrating the scholarly value of openness. Similarly, Tenopir et al. (2011) reported that while researchers recognize the importance of data sharing, barriers such as lack of infrastructure, unclear policies, and insufficient recognition hinder widespread adoption. These findings suggest that technical solutions alone are insufficient; institutional incentives and policy clarity are equally important.

2. Privacy and Ethical Risks in Data Sharing:

A major stream of literature focuses on privacy, confidentiality, and re-identification risks, particularly for sensitive or personal datasets. Ohm (2010) argues that even anonymized datasets can often be re-identified using auxiliary information, creating ethical and legal challenges for open data initiatives. In health research, Gymrek et al. (2013) demonstrated that individuals could be re-identified from supposedly de-identified genomic data through surname inference, highlighting technical vulnerabilities in open-data environments. These studies emphasize that privacy-preserving methods—such as encryption, controlled access, or differential privacy—are essential when archiving sensitive data.

3. Technical and Organizational Challenges in Secure Data Archiving:

Research also emphasizes the need for robust infrastructure and metadata standards to ensure long-term data preservation. Yakel and Faniel (2013) note that high-quality metadata and documentation are critical for enabling future reuse, yet they are often incomplete or inconsistently applied. Vines et al. (2014) showed that availability of research data decreases sharply over time, with data loss linked to poor archiving practices and inadequate institutional support. These findings indicate that secure data archiving requires ongoing stewardship, sustainable repository systems, and institution-wide data management policies.

Objectives of the Study:

- Examine secure archiving techniques for research data.
- Analyze frameworks and policies promoting open data sharing.
- Identify risks and challenges in data management.
- Propose strategies and best practices for balancing security and openness.
- Future research directions in data archiving and sharing

Research Methodology:

This study adopts a **qualitative-descriptive research design**, combining a review of existing literature, international policies, technical standards, and case studies related to secure archiving and open data sharing. The approach is **exploratory**, aiming to identify methods, frameworks, and challenges associated with balancing data security and openness.

The research draws upon:

1. **Primary sources**: Guidelines and recommendations from international organizations such as OECD, UNESCO, and the European Commission.
2. **Secondary sources**: Peer-reviewed articles, technical reports, and case studies from digital repositories and cyber security literature.

1) Secure Archiving: Principles, Infrastructure, and Methods:

Secure archiving refers to the long-term preservation of digital research data in a manner that ensures authenticity, integrity, confidentiality and continued usability. Effective archiving depends on a blend of technical solutions, standards, governance mechanisms, and institutional commitment.

1.1 Principles of Secure Digital Preservation:

1.1.1 Authenticity:

Authenticity ensures that data remains unaltered and trustworthy over time. Techniques such as cryptographic hashes, checksums, and version control systems are used to detect modifications. Audit trails and provenance metadata further establish a trusted chain of custody.

1.1.2 Integrity:

Integrity refers to the completeness and accuracy of data. Redundant storage, routine integrity checks, and controlled migration procedures help maintain high levels of data reliability.

1.1.3 Confidentiality:

Confidentiality is essential for sensitive datasets. Encryption, authentication systems, and policy-based access restrictions protect data from unauthorized access.

1.1.4 Usability:

A dataset must remain interpretable despite changes in technology. This requires comprehensive documentation, standardized formats, and proper metadata.These principles collectively support long-term preservation and ensure that archived data remains functional for future users.

1.2 Archival Frameworks and Standards:

The Open Archival Information System (OAIS) reference model (CCSDS, 2012) is the most widely recognized standard for digital archiving. It outlines six key functional components:

1. Ingest – data submission and metadata creation
2. Archival Storage – secure storage and redundancy
3. Data Management – indexing and searching
4. Administration – policy enforcement
5. Preservation Planning – ongoing system updates
6. Access – dissemination to authorized users

OAIS compliance ensures structured, interoperable, and sustainable archive operations.

1.3 Other relevant standards include:

ISO 14721 – OAIS implementation

ISO 16363 – trusted repository certification
PREMIS – metadata for digital preservation
Dublin Core – general metadata standards

2. Technological Methods for Secure Archiving:

2.1 Encryption:

Modern encryption protocols (AES-256, TLS 1.3) protect data from interception and unauthorized access. Encryption at rest protects stored data. Encryption in transit secures data during transfer.

2.2 Role-Based Access Control (RBAC):

Access is granted based on user roles and responsibilities. This reduces risks from internal misuse and human error. Combined with multi-factor authentication, RBAC forms a strong security backbone.

2.3 Redundant and Distributed Storage:

Strategies include:

LOCKSS networks ("Lots of Copies Keep Stuff Safe")
Cloud-based replication
Geographically distributed servers
These reduce the risk of catastrophic loss due to disasters or system failures.

2.4 File Format Migration and Emulation:

Digital obsolescence is a major threat. Migrating data to stable formats (e.g., CSV, PDF/A, TIFF) or using emulation to recreate old environments ensures long-term accessibility.

2.5 Automated Integrity Monitoring:

Systems routinely verify data integrity using checksum comparisons. Alerts are generated when files become corrupted or inaccessible.

3. Open Data Sharing: Frameworks, Philosophies, and Tools:

Open data sharing refers to making research data available to other researchers, institutions, and the public. While secure archiving focuses on preservation, open data prioritizes accessibility and reusability.

3.1 FAIR Principles in Open Data:

- The FAIR principles (Wilkinson et al., 2016) guide the creation of open datasets that remain useful across disciplines and technologies.
- Findable – persistent identifiers (e.g., DOIs) and searchable metadata
- Accessible – open protocols and clear licenses
- Interoperable – standardized formats (e.g., JSON, XML)
- Reusable – rich descriptions, provenance, and usage rights
- FAIR does not require full openness; it supports both open and controlled-access data.

3.2.1 Policy Frameworks for Open Science:

3.2.1 UNESCO Open Science Recommendation (2021)

Calls for broad global access to scientific data while maintaining ethical standards.

3.2.2 OECD Principles and Guidelines (2015):

Promotes open access to data from publicly funded research.

3.2.3 European Commission

Under Horizon Europe, open data is the default, with exceptions for privacy, security, and IPR. These policies collectively shape the global movement toward open science.

3.3.1 Open Data Repositories and Infrastructure

- Popular repositories include:
- Zenodo (EU OpenAIRE program)
- Dryad (life sciences)
- Figshare (multidisciplinary)
- ICPSR (social sciences)
- GenBank (genomics)

Repositories provide:

- Long-term archiving
- Persistent identifiers
- Licensing options (Creative Commons)
- Controlled-access features for sensitive datasets

3.3.2. Licensing and Legal Considerations

- Creative Commons licenses allow researchers to clarify reuse rights.
- Common licenses include:
- CC-BY (attribution required)
- CC-BY-NC (non-commercial use)
- CC0 (public domain)
- Legal frameworks such as GDPR (2016) and HIPAA govern data involving personal information.

4. Risks and Challenges in Secure Archiving and Open Sharing

While the benefits of open data are significant, risks can be equally substantial.

4.1 Privacy, Ethics, and Re-identification Risks

- Even anonymized datasets may be vulnerable.
- Narayanan & Shmatikov (2008) showed that cross-referencing datasets can reconstruct identities.
- Challenges include:
- Insufficient anonymization
- Sensitive health, genetic, or behavioral data
- Cultural or community-specific knowledge requiring protection
- Research involving Indigenous communities requires community-level consent and culturally appropriate protocols.

4.2 Cyber security Threats

Modern research infrastructure faces sophisticated attacks, including:

- Ransomware
- Malware
- Data tampering
- Unauthorized access through phishing
- Insider threats

Cloud-based repositories, despite being robust, still face risks if misconfigured or poorly monitored.

4.3 Data Misuse and Unintended Consequences

- Open data can be:
- Misinterpreted by non-experts
- Used for discriminatory purposes (e.g., predictive policing)
- Commercialized without proper attribution or compensation
- Taken out of cultural or scientific context
- Ethical governance is essential to mitigate misuse.

4.4 Sustainability and Infrastructure Limitations

- Data repositories require continuous funding for:
- Hardware storage
- Software updates
- Skilled personnel
- Data migration
- Preservation planning
- Many repositories struggle to maintain long-term sustainability.

4.5 Intellectual Property and Competitive Concerns

- Researchers may fear being “scooped” if data is shared before publication.
- Universities may restrict data dissemination to protect patents, commercial partnerships, or ongoing studies.
- Balancing openness with academic competitiveness remains a challenge.

5. Best Practices and Strategies for Balancing Openness and Security

5.1 Robust Data Management Plans (DMPs)

- Funding agencies increasingly require DMPs that include:
- Security protocols
- Storage plans
- Sharing timelines
- Ethical safeguards
- Metadata standards

5.2 Controlled Access Models

- Not all data should be fully open.
- Tiered models include:
- Open access
- Restricted access (application required)
- Highly controlled access (e.g., secure enclaves)

5.3 Advanced Anonymization and Privacy-Preserving Technologies

Examples:

- Differential privacy
- k-anonymity
- Secure multiparty computation
- Federated learning
- These reduce risk while enabling data analysis.

5.4 Regular Security Audits:

- Institutions should follow frameworks such as:
- NIST SP 800-53
- ISO/IEC 27001
- Routine vulnerability assessments help maintains resilience.

5.5 Education and Capacity Building

- Researchers must be trained in:
- Ethical data use
- Repository selection
- Harm reduction practices
- Legal responsibilities**.**

Conclusion:

Secure archiving and open data sharing are critical components of modern scientific ecosystems. When properly managed, open data enhances collaboration, accelerates discovery, and strengthens public trust in research. However, these benefits can only be realized through robust security measures, ethical governance, and sustainable infrastructure. A balanced approach rooted in FAIR principles, strong encryption, controlled access, privacy-preserving techniques, and long-term preservation standards enables institutions to harness openness while mitigating risks. As global science moves toward unprecedented levels of data sharing, the future of research will depend on systems that are both secure and open, supporting an equitable and innovative scientific community.

References:

1) CCSDS. (2012). Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems.

2) European Commission. (2016). General Data Protection Regulation (GDPR).

3) Gymrek, M., et al. (2013). Identifying personal genomes by surname inference. Science, 339(6117), 321–324.

4) Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. IEEE Symposium on Security and Privacy.

5) Narwade, M. R., & Jadhav, V. S. (2022). *Electronic Resource Management in Academic Libraries in India*. *International Journal of Classified Research Techniques & Advances*, 2(1), 26–?. ISSN 2583-1801.

6) NIST. (2020). Security and Privacy Controls for Information Systems and Organizations (NIST SP 800-53 Rev. 5). National Institute of Standards and Technology.

7) OECD. (2015). Making Open Science a Reality. OECD Publishing.

8) Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA Law Review, 57(6), 1701–1777.
9) Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. PeerJ, 1, e175.
10) Tenopir, C., et al. (2011). Data sharing by scientists: practices and perceptions. PLoS ONE, 6(6), e21101.
11) UNESCO. (2021). UNESCO Recommendation on Open Science.
12) Vines, T. H., et al. (2014). The availability of research data declines rapidly with article age. Current Biology, 24(1), 94–97.
13) Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(160018).
14) Yakel, E., & Faniel, I. M. (2013). Digital preservation: Metadata and curation. Annual Review of Information Science and Technology, 47(1), 295–335.