

*Type of the Paper: Article*

# A Reproducible Architecture for Automated Document Management with SharePoint Online and Microsoft Entra: A Design Study at PUCE Esmeraldas

Luis Adrián Fernández Tenelanda<sup>1\*</sup>, Danna Yolanda Solis Espinoza<sup>2</sup>

<sup>1</sup> Pontificia Universidad Católica del Ecuador Sede Esmeraldas

<sup>2</sup> Pontificia Universidad Católica del Ecuador Sede Esmeraldas

luis.fernandez@pucese.edu.ec

**Abstract:** University document management is often fragmented across folders, email, individual permissions, and approval paths with limited traceability. This paper presents and evaluates a reproducible architecture for an institutional document system using SharePoint Online, PnP PowerShell, and Microsoft Entra. Action design research examined scripts, configuration, and logs from March to August 2026. The artifact declares 18 hubs and 24 sites and, for each processed site, six libraries and 24 standard folders. Documented provisioning time fell from approximately 180 to 5 minutes (97.2%), while capacity increased from 10–15 to more than 150 sites per day. Nineteen CSV files provided 70 operational events from July 20 to August 5. The paper documents the architecture, BPMN workflow, automations, controls, evidenced licensing, and inference limits. Advanced Entra, Power Platform, and Purview capabilities without verifiable exports are reported as NR. The contribution is a replicable technical and methodological design for higher education, accompanied by a pseudonymized reproducibility package.

**Keywords:** Document management; SharePoint Online; Microsoft Entra; Automation; Higher education

## Introduction

Administrative documentation is a university infrastructure: it supports decisions, academic evidence, procurement, research, community engagement, human resources, and compliance. When a document is managed as an isolated file, the cost appears as search time, duplication, loss of context, delayed authorizations, and exposure of information. Digital transformation requires redesigning the value flow and responsibilities, not merely moving folders to the cloud [4,5]. The literature on digital transformation in higher education highlights that sustainable initiatives articulate strategy, processes, technology, governance, and human capabilities [1]. Within this framework, SharePoint Online provides libraries, metadata, search, versioning, and collaboration; Power Automate and Power Apps instantiate workflows and interfaces; Microsoft Entra constitutes the identity and access plane. The scientific problem consists of explaining how to assemble these components as a reproducible and evaluable organizational artifact, rather than describing a collection of configurations. The objective is to design and document an automated document management architecture applicable to the case of PUCE Esmeraldas. The questions are: (P1) how to translate document requirements into information architecture and controlled workflows? (P2) what automation pattern preserves traceability, portability, and least privilege?; and (P3) what evidence allows evaluating the artifact without revealing sensitive information? The novelty privileging, within a single design, the process map, taxonomy, script catalog, identity administration, and measurable operation.

## 2. Conceptual Framework

### 2.1 Document management and information architecture

ISO 15489-1 conceives records as evidence of activities and demands authenticity, reliability, integrity, and usability. In SharePoint, these attributes are operationalized through defined owners, content types, required metadata, version history, auditing, retention, and disposition rules. Microsoft recommends planning the architecture based on user needs and employing columns and content types as central elements to organize and retrieve information [11]. The design avoids two extremes: a deep folder hierarchy, which hides context and hinders analysis, and a single unclassified library, which concentrates risk and degrades the user's experience. The unit of design is the document family, associated with a purpose, owner, metadata schema, access policy, retention, and approval workflow.

### 2.2 Action Design Research

Design science research places the artifact and its evaluation at the center of applied knowledge [2]. Action design research adds that the artifact is configured during its interaction with the organizational context, combining construction, intervention, and evaluation [3]. This perspective is relevant because permissions, taxonomies, and workflows change when confronted with real cases, exceptions, and institutional responsibilities. The studied artifact comprises five inseparable elements: information model, technical architecture, automation catalog, identity model, and operation package. Its quality is judged by utility, control, maintainability, and the ability to produce verifiable evidence.

## Methodology

A single instrumental case study based on action design research is proposed. The unit of analysis is the institutional document system and its processes of reception, classification, review, approval, publication, consultation, retention, and disposition. The corpus must gather internal regulations, inventory, diagrams, exported configuration, script versions, execution logs, incidents, and interviews with process owners. For this version, the available corpus is limited to the general topic, institutional identity, and the description of the implementation; therefore, figures and tenant identifiers remain as validation fields. The procedure comprises four cycles: (1) problem formulation and selection of document families; (2) construction and intervention through prototypes; (3) concurrent evaluation with functional, security, and usability testing; and (4) formalization of design principles. Triangulation combines configuration inspection, telemetry, audit logs, and user perception. No secrets, personal identifiers, or sensitive paths should be incorporated into the publishable dataset. The rigor criteria are requirement-configuration-test traceability, separation between environments, version control, decision logging, negative cases, and reproducibility through declarative infrastructure. Quantitative evaluation prior to submission must report period, number of families, documents, eligible users, workflows, executions, and failures, along with confidence intervals where appropriate.

Table 1. Operationalization of constructs, evaluation indicators, and acceptance criteria

| Construct       | Observable Indicator                    | Source of Evidence                              | Acceptance Criteria                                        |
|-----------------|-----------------------------------------|-------------------------------------------------|------------------------------------------------------------|
| Traceability    | Record-state-approvals and relationship | Versioning and log                              | Target: 100%; record result NR                             |
| Integrity       | Versions, hash, or equivalent control   | Library and validation script                   | Planned counts and reconciliation; final sample NR         |
| Security        | Group permissions and exceptions        | Entra, SharePoint, and access review            | Permissions are not copied by default; total review NR     |
| Portability     | Hardcoded parameters isolated from code | Solution, environment variables, and repository | DEV-TEST-PROD deployment without manual workflow editing   |
| Maintainability | Detected and recoverable failures       | Telemetry, alerts, and runbook                  | Every automation has an owner, SLA, and exception handling |

The quantitative provisioning intervention was observed from March 4 to August 5, 2026 (155 calendar days). The migration log sub corpus covers July 20 to August 5, 2026, inclusive (17 days). For the governance analysis, the document review was extended through August 22, 2026 (172 days from the start). These windows are declared separately to avoid attributing metrics captured solely in the July–August logs to the entire project. Source traceability and pseudonymization were applied. The publishable corpus contains no administrative emails, application identifiers, usernames, secrets, private paths, or Tenant object identifiers. Data that do not appear in verifiable exports or logs are marked as non-reported

(NR), rather than imputed.

Table 2. Primary evidence corpus, source artifacts, and analytical application

| ID  | Verified Evidence                                                   | Period / Version                       | Scientific Use                                              |
|-----|---------------------------------------------------------------------|----------------------------------------|-------------------------------------------------------------|
| E01 | Master provisioning script (950 lines; SHA-256 AAE76A1261608DA2...) | Aug 5, 2026                            | Architecture, objects, and implemented functions            |
| E02 | Migration engine (698 lines; SHA-256 BE078FBFA5F9A5CC...)           | Jul 30, 2026                           | Simulation, execution, resumption, and reconciliation       |
| E03 | Technical README and local configuration                            | PowerShell 7.6.4; PnP.PowerShell 3.1.0 | Requirements, parameters, and operational controls          |
| E04 | 19 CSV files with 70 events                                         | Jul 20 to Aug 5, 2026                  | State distribution and migration scope                      |
| E05 | Provisioning measurement log                                        | Mar 4 to Aug 5, 2026                   | 180→5 min and 10–15→>150 sites/day                          |
| E06 | Component replication and hub association evidence                  | 2026                                   | Pages, navigation, libraries, lists, backup, and validation |
| E07 | Interactive authentication log with Entra registered application    | 2026                                   | Workload identity without embedded secrets                  |
| E08 | Technical history of simulation, confirmation, and restoration      | 2026                                   | WhatIf, explicit confirmation, backup, and logical rollback |

## Results: Implementation of the Document Management Artifact

Table 3. Baseline and observed results

| Metric                     | Baseline             | Result                            | Change / Interpretation                                 | Source |
|----------------------------|----------------------|-----------------------------------|---------------------------------------------------------|--------|
| Provisioning time per site | ≈180 min             | ≈5 min                            | −175 min; 97.2% reduction                               | E05    |
| Documented daily capacity  | 10–15 sites/day      | >150 sites/day                    | >10× compared to manual upper bound                     | E05    |
| Declared structure         | Unnormalized         | 18 hubs + 24 sites                | 42 governable nodes in the manifest                     | E01    |
| Document pattern per site  | Heterogeneous        | 6 libraries + 24 standard folders | Repeatable and idempotent structure                     | E01    |
| Migration logs             | No comparable corpus | 19 CSVs; 70 events                | 37 SIMULATED, 19 COPIED, and 14 ERROR                   | E04    |
| Registered migration scope | NR                   | 4 sources; 5 destinations         | Observed coverage; does not represent the entire Tenant | E04    |

As observed in Table 3, the reduction was calculated as  $(180 - 5) / 180 \times 100 = 97.2\%$ . The capacity increase is expressed conservatively as greater than tenfold because >150 is compared with the upper bound of the baseline (15). The 14 ERROR logs do not constitute a failure rate: the log mixes simulations, retries, and heterogeneous events, lacking a denominator of independent executions. Declaring this limitation avoids ungrounded causal inference. The master script declares six libraries per site—administrative management, processes and procedures, regulations, document management, tracking and control, and archive—with 24 standard folders. The architecture separates provisioning, migration, component replication, inventory, validation, and logging; personal permissions are not copied by default.

Table 4. Licensing and Tenant controls

| Service / Control                            | Status in Case                                                 | Evidence                                                  | Editorial Treatment                                      |
|----------------------------------------------|----------------------------------------------------------------|-----------------------------------------------------------|----------------------------------------------------------|
| SharePoint Online                            | Operational                                                    | Sites, hubs, libraries, and logs E01–E06                  | Can be stated as implemented                             |
| Microsoft Entra & PnP Application            | Operational                                                    | Registered application and interactive authentication E07 | Publish without IDs or accounts                          |
| MFA                                          | Reported in corpus                                             | Institutional authentication; policy export not included  | Limited assertion; do not attribute full coverage        |
| PowerShell / PnP.PowerShell                  | Operational; tools without standalone M365 SKU                 | Local versions E03                                        | Declare versions                                         |
| Power Automate, Power Apps & Power BI        | Designed; production execution not demonstrated in this corpus | No solution exports or telemetry                          | NR as implementation                                     |
| Conditional Access                           | Not evidenced                                                  | No policy exports or SKU                                  | Do not assert implementation; verify Entra ID P1 license |
| PIM, Access Reviews & Lifecycle Workflows    | Not evidenced                                                  | No configuration or SKU                                   | Treat as roadmap; verify P2/Governance                   |
| Microsoft Purview: Retention and Disposition | Not evidenced                                                  | No exported policies or labels                            | Treat as roadmap and validate plan                       |

In Table 4, it is indicated that the absence of evidence does not equate to the absence of the service: it means that the available corpus does not allow for its verification. Prior to submission, a licensing administrator must export the SKU, enabled units, and aggregated assignments from the admin center, and an administrator for each service must export the relevant configurations. The manuscript retains NR until those artifacts become available. In Figure 1, the resulting architecture separates experience, processes, information, identity, and observability. This separation reduces dependencies: the interface can change without altering the document model; workflows consume parameters and connection references; and Entra groups govern access without hardcoding individuals into scripts. The compliance layer is applied through classification, retention, and auditing, subject to licensing and institutional policy. The pattern favors purpose-driven sites and libraries with stable permission inheritance and documented exceptions. Content types encapsulate shared metadata and behavior. Views respond to specific tasks—pending items, by department, by status, upcoming expirations—rather than replicating physical folder structures. Search depends on metadata quality and managed properties; therefore, completeness is measured starting from the pilot phase.

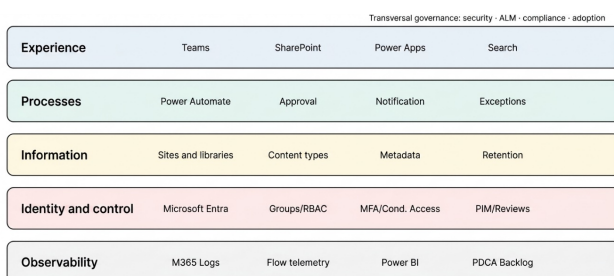


Figure 1. Logical architecture of the document management system

Gateway prevents executing live changes without manifest validation. Unapproved branches return to correction; execution exceptions trigger backup and logical rollback. Authentication uses a registered application and an interactive institutional account; the diagram does not publish Tenant identifiers Figure 2.

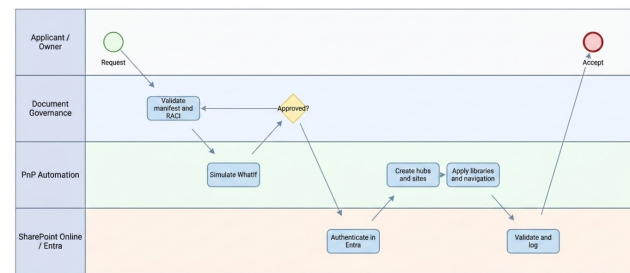


Figure 2. Simplified BPMN map of provisioning and acceptance

#### 4.1. Implementation sequence in twelve steps

1. Define scope, sponsorship, and document families. Identify the target service receiving value, process owner, regulatory constraints, and success criteria.
2. Conduct inventory and classify risk. Record source, volume, format, duplicate data, sensitivity, retention, permissions, and metadata quality; do not migrate ownerless content.
3. Map the current state. Measure transfers, wait times, rework, approvals, and loss points; select a high-value pilot with manageable risk.
4. Design the taxonomy. Establish controlled vocabulary, site columns, content types, record identifiers, and mandatory field rules; minimize manual data capture required from users.
5. Design sites, libraries, and lifecycle. Configure owners, versioning, approvals, views, indexing, retention, and disposition; justify any permission inheritance breaks.
6. Model identity in Microsoft Entra. Create role- or function-based groups, separate administrative from operational functions, enforce least privilege, and document emergency exclusions.
7. Register workload identities. Prefer managed identities or certificates over secret credentials; scope Microsoft Graph permissions and, where feasible, restrict site-level access.
8. Build solutions-bound automations. Utilize connection references, environment variables, consistent naming conventions, exception handling, bounded retries, and execution correlation IDs.
9. Automate provisioning. Declare sites, columns, content types, libraries, views, groups, and settings via PnP PowerShell or Microsoft Graph; ensure every script is idempotent.
10. Migrate and reconcile. Execute small batches, log source and target endpoints, validate counts, sizes,

metadata, and permissions, and maintain a log of actionable errors.

**11. Pilot, train, and accept.** Validate primary user flows, edge cases, access denials, accessibility, and recovery procedures; conduct task-based training and track active adoption rather than attendance alone.

**12. Operate and continuously improve.** Publish operational runbooks, designated owners, maintenance windows, KPI dashboards, periodic access reviews, and a PDCA backlog; decommission legacy systems only after verifying operational continuity.

#### 4.2 Automation and script catalog

Table 5. Catalog of Automations and Scripts for the Document Management System

| ID     | Automation                                                   | Technology                    | Input / Output                          | Control and Telemetry                                         |
|--------|--------------------------------------------------------------|-------------------------------|-----------------------------------------|---------------------------------------------------------------|
| AUT-01 | Provisioning of sites, columns, content types, and libraries | PnP PowerShell / Graph        | YAML/JSON manifest → configuration      | Idempotency, per-object logging, logical rollback             |
| AUT-02 | Onboarding and review of role-based groups                   | Entra / Graph                 | RACI matrix → groups and members        | Ownership, expiration, review, and exception handling         |
| AUT-03 | Document registration and classification                     | Power Apps / SharePoint       | Form → metadata-enriched document       | Validation, unique ID, actor registration                     |
| AUT-04 | Review and approval                                          | Power Automate                | Initial state → decision and evidence   | Timeout, reassignment, reasoned rejection, correlation        |
| AUT-05 | Notification and escalation                                  | Power Automate / Teams        | Event/SLA → alert                       | Loop prevention, frequency control, and channel ownership     |
| AUT-06 | Batch migration                                              | PnP PowerShell                | Source → library and metadata           | Item count, file size, errors, resumption, and reconciliation |
| AUT-07 | Retention policy application                                 | Purview / SharePoint          | Class and event → label                 | Licensing, validation testing, authorized disposition         |
| AUT-08 | Telemetry extraction                                         | Graph / M365 Audit / Power BI | Events → analytical model               | Minimization, pseudonymization, and retention                 |
| AUT-09 | Detection of atypical permissions                            | Graph / PnP                   | ACL → findings                          | Allowlist, severity ranking, closure evidence                 |
| AUT-10 | Cross-environment deployment                                 | Solutions / Pipelines         | Package + settings → target environment | Approval, versioning, smoke testing, and rollback             |

Algorithm 1. Idempotent provisioning pattern (reference snippet)

```
$cfig = Get-Content $ManifestPath -Raw | ConvertFrom-Json
Connect-PnPOnline -Url $cfig.siteUrl -Interactive
foreach ($field in $cfig.fields) { if (-not (Get-PnPField -Identity $field.internalName -ErrorAction SilentlyContinue)) { Add-PnPField -DisplayName $field.displayName -InternalName $field.internalName -Type $field.type -Group $cfig.fieldGroup } }
# Repetir el patrón: consultar estado, comparar con el manifiesto,
# crear o actualizar, registrar resultado y nunca incrustar secretos.
```

```
$cfig = Get-Content $ManifestPath -Raw | ConvertFrom-Json
Connect-PnPOnline -Url $cfig.siteUrl -Interactive
foreach ($field in $cfig.fields) { if (-not (Get-PnPField -Identity $field.internalName -ErrorAction SilentlyContinue)) { Add-PnPField -DisplayName $field.displayName -InternalName $field.internalName -Type $field.type -Group $cfig.fieldGroup } }
# Repetir el patrón: consultar estado, comparar con el manifiesto,
# crear o actualizar, registrar resultado y nunca incrustar secretos.
```

Algorithm 1 illustrates the check–compare–converge principle. The institutional repository must incorporate version control, testing, static analysis, environment-specific parameters, structured logging, and peer review. Global permissions should not be granted if the task can be resolved with a narrower scope. Environment variables and connection references allow solutions to be transported without modifying their underlying logic, a central requirement for lifecycle management [9]. In Entra, conditional access policies must be deployed first in report-only mode with emergency access accounts excluded and closely monitored. Multifactor authentication, least privilege, temporary role activation, and periodic access reviews reduce standing access; their precise implementation depends on licensing and operational maturity [10,12].

#### 4.3 Testing and acceptance criteria

Table 6. Testing Matrix and Acceptance Criteria for the Document Management System

| Test Category | Test Case                                                    | Evidence                  | Expected Result                                             |
|---------------|--------------------------------------------------------------|---------------------------|-------------------------------------------------------------|
| Functional    | Create, review, reject, correct, and approve                 | History and notifications | Complete sequence without loss of metadata                  |
| Negative      | Unauthorized user attempts to access or approve              | Access log                | Denial of access and no data leakage via direct links       |
| Resilience    | Connector temporarily unavailable                            | Retries and error queue   | Recovery without duplicating actions                        |
| Migration     | Stratified sample by format and sensitivity                  | Source–target report      | Reconciled item count, file size, metadata, and permissions |
| ALM           | Import solution to TEST environment with modified parameters | Deployment log            | Enabled flows with valid connection references              |
| Recovery      | Restore version and revoke access                            | Exercise logbook          | Compliance with institutional RTO/RPO targets               |

## Discussion

The case confirms that the primary difficulty lies not in creating a library, but in aligning documentary meaning, authority, and operations. Information architecture determines what can be searched, measured, and retained; Entra determines who can act; and automation transforms rules into observable behavior. This interdependence explains why purely technical initiatives stall when process ownership is lacking. The contribution relative to general digital transformation studies lies in providing an implementable level of specificity. The pattern links every decision with evidence and prevents Power Automate from turning into a proliferation of personal flows. It also distinguishes portability from simple copying: a flow is portable when its connections, parameters, permissions, and dependencies are governed, not merely when it can be manually exported. The design has limitations. Microsoft 365 is an evolving service; licensing, operational limits, and feature names must be verified prior to production. Furthermore, a single case study does not allow universal causal claims. Future publication must incorporate the baseline, observation period, intervened document families, execution logs, and user perception metrics. Without this data, the findings should be interpreted as design validation rather than an estimation of causal effect.

## Conclusions

The implementation of an automated document management system in higher education requires sociotechnical architecture: taxonomy and lifecycle, portable automations, least-privilege identities, operational evidence, and continuous improvement. The twelve-step pattern provides a replicable roadmap to transition from an inventory to a governed service. Three core principles summarize the lessons learned: designing by document families and value, managing access by roles rather than individual accounts, and treating scripts and flows as products with ownership, versioning, testing, and telemetry. The next phase of research must measure effects on cycle time, rework, completeness, adoption, and risk, as well as publish an anonymized package of instruments and reproducible code.

### *Declarations Data and Code Availability*

The local package Repositorio\_reproducibilidad\_PUCESE.zip was prepared with metrics, evidence inventory, licensing, CRediT taxonomy, and pseudonymized BPMN diagrams. The raw Tenant corpus remains restricted. For the version of record, the same package must be deposited in Zenodo, OSF, or the institutional repository to obtain a persistent identifier (DOI).

### *Author Contributions (CRediT)*

Luis Adrian Fernandez Tenelanda: Conceptualization, Methodology, Software, Investigation, Data Curation, Validation, Visualization, Writing – Original Draft, and Project Administration. Danna Yolanda Solis Espinoza: Investigation, Resources, Formal Analysis, Validation, and Writing – Review & Editing. This statement must be ratified by both authors upon manuscript submission.

### *Funding*

No specific external funding was identified in the reviewed corpus; the implementation was carried out using institutional resources and infrastructure. This statement must be ratified by the authors.

### *Conflicts of Interest*

No conflicts of interest were identified in the reviewed corpus; the authors must confirm this statement in the editorial submission form.

### *Ethics and Privacy*

The analysis relies strictly on necessary operational metadata, adhering to data minimization, pseudonymization, and institutional approval where applicable. No personal names, sensitive file paths, object identifiers, or Tenant secrets may be published.

## *Use of AI Tools*

This version received support from a generative AI tool for structuring and editorial refinement. The authors retain full responsibility for the accuracy, originality, references, code, data, and conclusions, and must adapt this declaration to the specific policy of the target journal.

## References

1. L. M. C. Benavides, J. A. Tamayo Arias, M. D. Arango Serna, J. W. Branch Bedoya, and D. Burgos, "Digital transformation in higher education institutions: A systematic literature review," *Sensors*, vol. 20, no. 11, p. 3291, Jun. 2020, doi: 10.3390/s20113291.
2. A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, Mar. 2004, doi: 10.2307/25148625.
3. M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren, "Action design research," *MIS Quarterly*, vol. 35, no. 1, pp. 37–56, Mar. 2011, doi: 10.2307/23043488.
4. G. Vial, "Understanding digital transformation: A review and a research agenda," *The Journal of Strategic Information Systems*, vol. 28, no. 2, pp. 118–144, Jun. 2019, doi:10.1016/j.jsis.2019.01.003.
5. P. C. Verhoef et al, "Digital transformation: A multidisciplinary reflection and research agenda," *Journal of Business Research*, vol. 122, pp. 889–901, Jan. 2021, doi:10.1016/j.jbusres.2019.09.022.
6. Information and documentation—Records management—Part 1: Concepts and principles, ISO Standard 15489-1:2016, International Organization for Standardization, 2016.
7. Information security, cybersecurity and privacy protection—Information security management systems—Requirements, ISO/IEC Standard 27001:2022, International Organization for Standardization, 2022.
8. Microsoft, "Use retention labels to manage the lifecycle of documents stored in SharePoint," Microsoft Learn, 2025. [Online]. Available: [<https://learn.microsoft.com/en-us/microsoft-365/compliance/auto-apply-retention-labels-scenario>]
9. Microsoft, "Environment variables for Power Platform overview," Microsoft Learn, 2025. [Online]. Available: [<https://learn.microsoft.com/en-us/power-apps/maker/data-platform/environmentvariables>]
10. Microsoft, "The principle of least privilege with Microsoft Entra ID Governance," Microsoft Learn, 2025. [Online]. Available: [<https://learn.microsoft.com/en-us/entra/id-governance/scenarios/least-privileged>]
11. Microsoft, "Introduction to SharePoint information architecture," Microsoft Learn, 2026. [Online]. Available: [<https://learn.microsoft.com/en-au/sharepoint/information-architecture-modern-experience>]
12. Microsoft, "What is Conditional Access?" Microsoft Learn, 2026. [Online]. Available: [<https://learn.microsoft.com/en-us/entra/identity/conditional-access/overview>]
13. Microsoft, "DriveItem: delta," Microsoft Graph documentation, 2026. [Online]. Available: [<https://learn.microsoft.com/en-us/graph/api/driveitemdelta>]
14. Microsoft, "Configure identity and access management for Power Platform," Microsoft Learn, 2026. [Online]. Available: [<https://learn.microsoft.com/en-us/power-platform/guidance/adoption/conditional-access>]
15. National Institute of Standards and Technology, "Zero Trust Architecture," NIST Special Publication 800-207, Aug. 2020, doi: 10.6028/NIST.SP.800-207.
16. SciELO, "Program, publication model, and SciELO Network." [Online]. Available: [<https://www.scielo.org/en/about-scielo/program-publication-model-and-scielo-network/>]