Appendix 1: LIST OF DOCUMENTS NEEDED TO SUPPORT APPLICATION

The following is a recommended list of supporting documents to facilitate completion of the tool and to provide as part of your NEST Mark: Real-World Data Source Tool submission, *as applicable to your data source.*

Section 1: General Data Holder and Data Source Information

- Priority
  - ☐ Standard "data source description" (what it is; how it's captured; known limitations)
  - ☐ Prior RWD use/credibility: list of notable publications and/or regulatory use cases (if any)
  - ☐ Documentation on how medical devices are identified in the data source

Section 2: Governance and Quality Management System

- Priority
  - ☐ Governance overview (including organizational chart, roles [including data stewardship roles] and responsibilities and escalation)
  - ☐ QMS/quality program overview (or QA program description) and Corrective and Preventive Action (CAPA) processes
  - ☐ SOP index or controlled document list
  - ☐ SOP on SOPs (*process for developing, approving, updating, and training on SOPs)*
  - ☐ Change control policy / Versioning SOP and change log
  - ☐ Incident/problem management policy with sample redacted incident ticket or postmortem template
  - ☐ Subprocessor inventory and functions (cloud/ETL/de-ID/linkage/analytics)
  - ☐ Training and competency program (beyond SOPs)
  - ☐ Regulatory compliance SOP *(audits and inspections)*
- Preferred
  - ☐ Internal audit process summary with example of an audit finding and resolution (redacted)
  - ☐ Sample external audit and/or inspection reports
  - ☐ Formal RACI for data curation and release sign-offs
  - ☐ Research Integrity and research methodologies SOP (as applicable)

Section 3: Data consistency & Stability and Quality Control Framework
Section 4: Data Accrual, Traceability, and System and Data Source Versioning
Section 5: ETL and Technical Controls
Section 6: Data Integration & Linkage

- Priority
  - ☐ Data accrual / capture SOP and documentation (including data entry guidelines, edit checks, coding guidelines, automatic data feeds, etc.)
  - ☐ Data provenance / traceability SOP and reports from original source to "research-ready" database
  - ☐ Data model/schema documentation + versioning approach
  - ☐ Data dictionary + coding/terminology mappings (ICD, CPT/HCPCS, LOINC, etc.)
  - ☐ ETL architecture overview (including diagrams)
  - ☐ ETL and integrity/completeness checks SOP (Extraction, loading, transformation. Include validation of transformations and methods to demonstrate completeness, with respect to the data dictionary)
  - ☐ Source-to-target mappings / lineage documentation
  - ☐ Data quality control (QC) and accuracy SOP (*edit checks, thresholds, verification, escalation steps used)*
  - ☐ Data quality assurance (QA) SOP (*Please provide with respect to the data dictionary*)
  - ☐ Sample QC/QA report(s) and evidence of issue closure (redacted)

☐ Logging/monitoring and audit log retention policy

☐ Audit trails SOP

☐ Data linkage SOP / process documentation (within the data source and/or with other data sources) and reports

☐ Software validation SOP (Software development life cycle (SDLC) and system validation)

- Preferred
  - ☐ Automated lineage tool output (if available)
  - ☐ Formal validation test scripts/checklists used for releases

Section 7: Data Access & Sharing, Privacy & Security, and Regulatory Transparency

- Priority
  - ☐ Data provenance and rights statement (who provides the data; authority to use/share)
  - ☐ Template DUA/BAA (redacted acceptable) and data use policy
  - ☐ Access request/approval workflow (how projects are approved; criteria)
  - ☐ Retention and destruction policy
  - ☐ De-identification methodologies and applicability SOP (Safe Harbor and/or Expert Determination summary)
  - ☐ Re-identification risk management approach (thresholds/mitigations)
  - ☐ Tokenization/pseudonymization approach (if used)
  - ☐ Disclosure control rules for outputs (e.g., min cell size/suppression policy)
  - ☐ System and data security program overview (SOP/policies/governance)
  - ☐ Independent assurance: SOC 2 Type II (preferred) or ISO 27001 (or equivalent) with scope statement
  - ☐ Access control SOP and policies (RBAC, provisioning/deprovisioning, privileged access)
  - ☐ MFA/SSO policy
  - ☐ Encryption standards and key management overview
  - ☐ Vulnerability management/patching policy
  - ☐ BCP/DR plan with RTO/RPO targets
  - ☐ IRB/ethics posture summary (if applicable)
- Preferred
  - ☐ Standard language for onward transfer and/or re-disclosure restrictions
  - ☐ Privacy impact assessment template or equivalent
  - ☐ Evidence of periodic re-assessment of de-identification risk
  - ☐ Penetration test executive summary + remediation tracking
  - ☐ Evidence of disaster recovery testing (annual test summary)

Section 8: Data Source Representativeness, Continuity of Care, and Longitudinality
Section 9: Quantitative Data Source Characteristics

- Priority
  - ☐ Coverage profile: sites/geography/settings/time span; refresh cadence/latency
  - ☐ Longitudinality/continuity metrics: follow-up distributions; encounter density
  - ☐ Missingness profiles for key fields (by site/time where feasible)
  - ☐ Representativeness assessment (benchmarks if available)
  - ☐ Linkage methodology + validation metrics (if linkage is offered)