Who's Afraid of a Digital TMF? Challenges and Opportunities in the Future of the TMF Standard

Presented by Jay Smith, Vice President Product, TransPerfect Life Sciences

Meet the Speaker

Jay Smith

Title: Vice President Product, Trial Interactive

Organization: Transperfect Life Sciences

Jay currently leads product management at TransPerfect, where he is responsible for the Trial Interactive platform. Prior to joining TransPerfect, he led product teams at Medidata Solutions, Sparta Systems, VenueNext, and Cureatr.

Over his career, Jay has overseen the creation and management of solutions spanning eTMF, CTMS, EDMS, QMS, RTSM, EDC, LMS, and eISF. Earlier in his career, he also developed and managed products for RIMS, RDMS, and Submission Publishing.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

Agenda

1. TMF Challenges
2. Industry & Strategic Drivers
3. Digital Data Flow
4. Digital Clinical Ecosystem
5. eTMF Interconnects
6. Digital Document Workflows
7. AI & Hyperautomation
8. Benefits & Learnings
9. Benefits Gained
10. Roadmap
11. Next Steps

Introduction

Current state of trial master file (TMF):

- Still heavily document-centric
- Historically composed of documents, PDFs, and scanned signatures
- Metadata still limited, manual, and inconsistent
- Metadata definition varies greatly
- Wildly different coding and metadata standards with a shifting landscape
- Systems largely disconnected

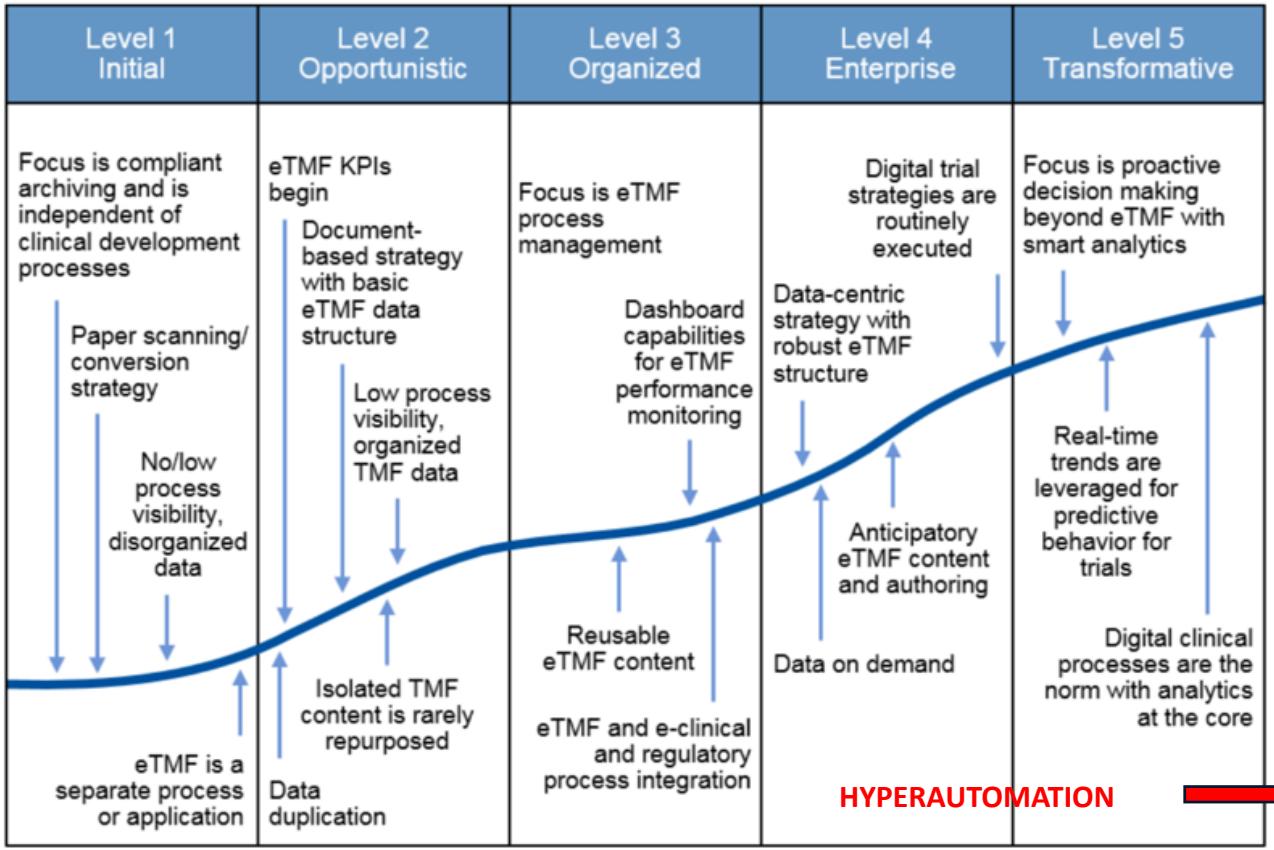
Challenges with the Current Model

- High manual effort
- Poor interoperability
- TMF treated as only an archive
- Monolithic silos
- Standard is permissive, de facto
- Hard to leverage for analytics or AI

Industry Drivers

- eTMF systems are becoming more advanced, incorporating data structure, CTMS, and complementary systems
- Agency harmonizations like EU CTR (EU 536/2014) and FDAAA 801 will simplify registration and study start up in different regions
- Larger sites starting to leverage their own enterprise site technology instead of systems provided by sponsors and CROs
- IRB/ECs are (slowly) undergoing digital transformations
- Multi-modal data fusion frameworks, often using AI, are getting a foothold to simplify EDC
- Use of decentralized trials is accelerating with more patient centricity
- TransCelerate and other vendor-based platforms making steady progress, for example eConsent
- Adoption of ICH M11, ADaM, FHIR, ODM, eCTD continue to grow

eTMF Strategic Drivers

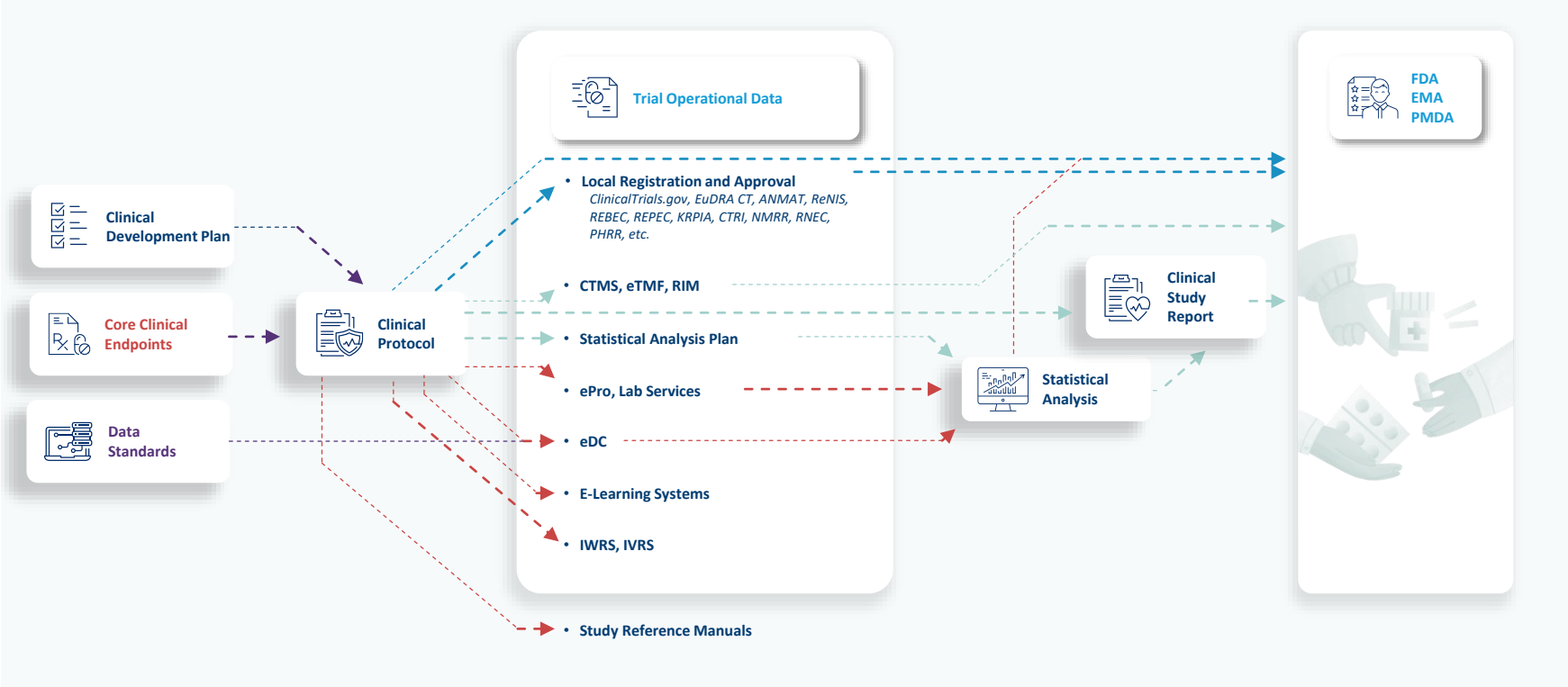


How can we...

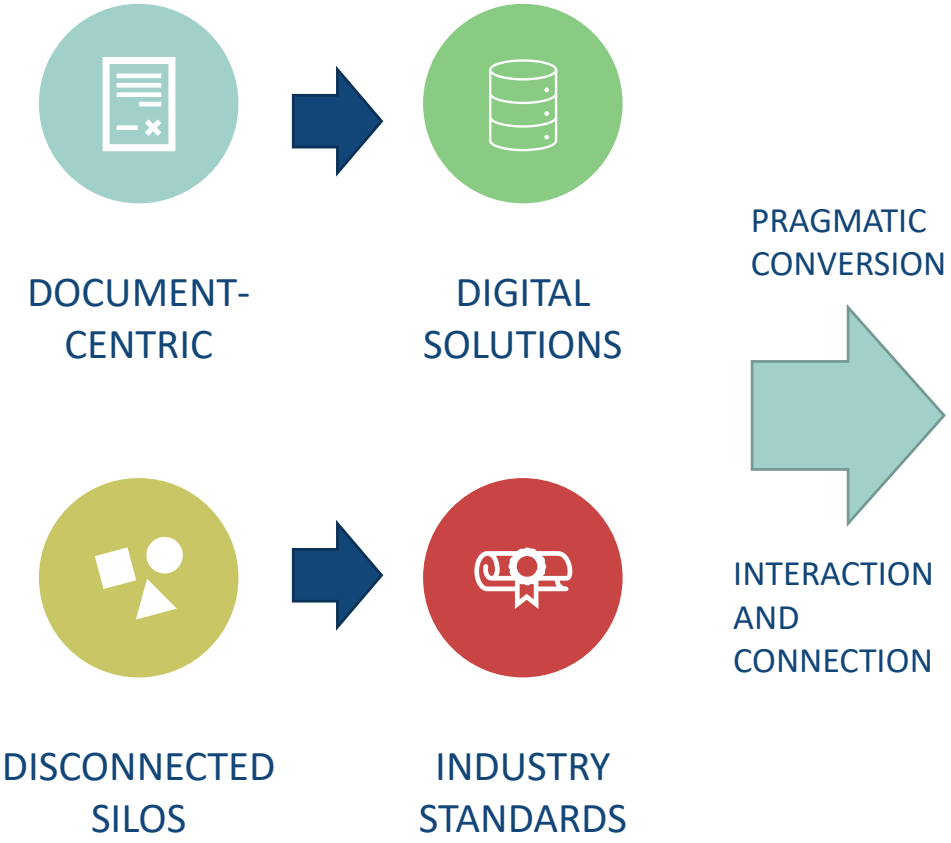
- Transform workstreams?
- Capture real metadata?
- Make proactive decisions?
- Interoperate digitally?

Source: Gartner, "The Trial Master File Function Lifecycle: From Reactive to Strategic," 2019

Clinical Trial Information Flow



The Continuing Journey



The Digital Clinical Ecosystem

Electronic Data Capture (EDC)

Clinical Trial Management System (CTMS)

Sites and Electronic Investigator Site File (eISF)

Learning Management System (LMS)

Institutional Review Board (IRB/EC) Regulators and Agencies

Electronic Trial Master File (eTMF)

Randomization and Trial Supply Management (RTSM)

Clinical Data Repository and Analytics

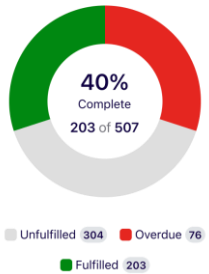
eTMF Inter-Connects

Business Goals for eTMF:

- eTMF health is timely and contemporaneous
- Completeness and accuracy
- Quality and inspection-ready

Ecosystem Connects:

- Sites, organizations, and contacts
- Regulators, External parties, third parties, and affiliates
- Business arrangements constantly shifting



Clinical Trial Awareness:

- Data flows exist between EDC, IXRS, RBQM, and CTMS
- eTMF must be "trial aware" of study milestones, visits, events, and status

Document Data Flows:

- Clinical, quality, and RIM document workflows can impact the eTMF
- eTMF should be aware of the data classifications

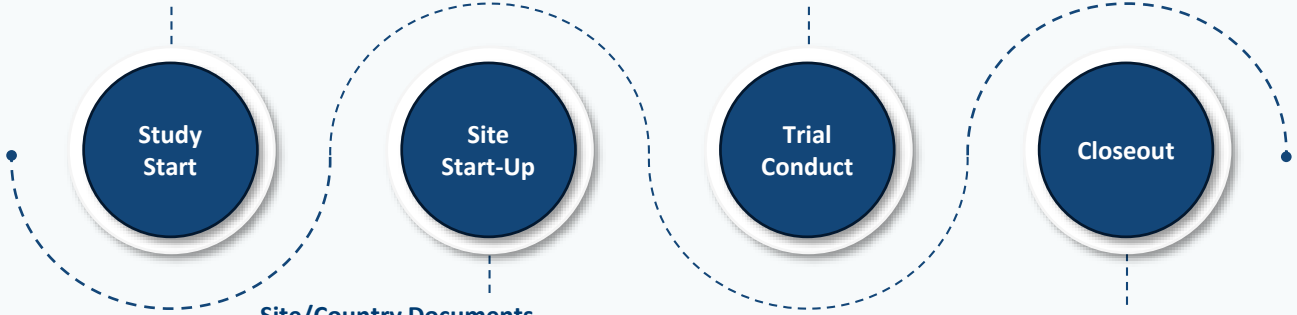
eTMF Integration Points

Protocol Creation

- Prepare eTMF with basic information needed to provision the study (IP, protocol #, etc.)
- Trial information - product, protocol number, protocol title, phase, type
- Labs and third-party vendors

Mid-Trial Updates

- New sites indicates new sets of docs against site milestones
- New investigator requires new 1572, CV, etc.
- Monitoring approved visit reports and letters
- Study, country, and site milestones that indicate a new document must be provided (e.g., protocol amendment)



Site/Country Documents

- Country/study status based on TMF milestone completion
- Countries - where the study will be conducted
- Automatic update of clinical sites that will participate and the principal investigators, sub-investigators, and site staff who will participate

Study Milestone Updates

- DB lock and archival milestones and study status that close out and archive the trial in the TMF
- Final regulatory documents archived in the TMF

eTMF Integration Points

- Studies
- Countries
- Sites
- Investigators
- Contacts
- Participants
- Visits
- Milestones
- Safety/PV
- Issues
- Activities
- Organizations
- Events and Milestones
- Documents

Digitization of Trial Documents into Fully Digital Business Workflows

1572 Form	Electronic form filled out that provides investigator, site, contact, and organization data to CTMS and eTMF
Delegation Log	Written form that tracks site personnel changes with their delegations, used for access, training, and documentation requirements.
SUSAR	AE/SAE forms with fielded data is extracted into document metadata, notification workflow is initiated to capture acknowledgements from investigators and agencies are notified
Informed Consent	eConsent process is initiated that activates a patient account, provides remote telephony or virtual meeting to ensure comprehension and compliance, and captures an eSignature (eIC)
Training Certificate & Evidence	Training Certificates and Logs are outputs of the Site Training program, Investigator Meetings, eLearning, coursework, and compliance activities. These must be stored in the eTMF, recorded in the CTMS, and carefully tracked.
Clinical Trial Agreement	Can be generated from the budgeting system and reviewed online in a collaboration space, or directly from a contracting solution
Feasibility Questionnaire	Fully-electronic feasibility questionnaires can provide consistent data from sites that may be captured and re-used to assist in planning and future decision-making
eCRF	EDC, ePRO, eCOA
Protocol Deviation	Electronic form that can be captured immediately into a workflow in the CTMS
Note to File	A Quality Management workflow, with QA signatories and approvers
Subject Enrollment Log (Redact)	Can be generated from the EDC and generated (if necessary) to an electronic document as evidence
Site Monitoring Log	Can be generated from the EDC and generated (if necessary) to an electronic document as evidence

Metadata is the New Index

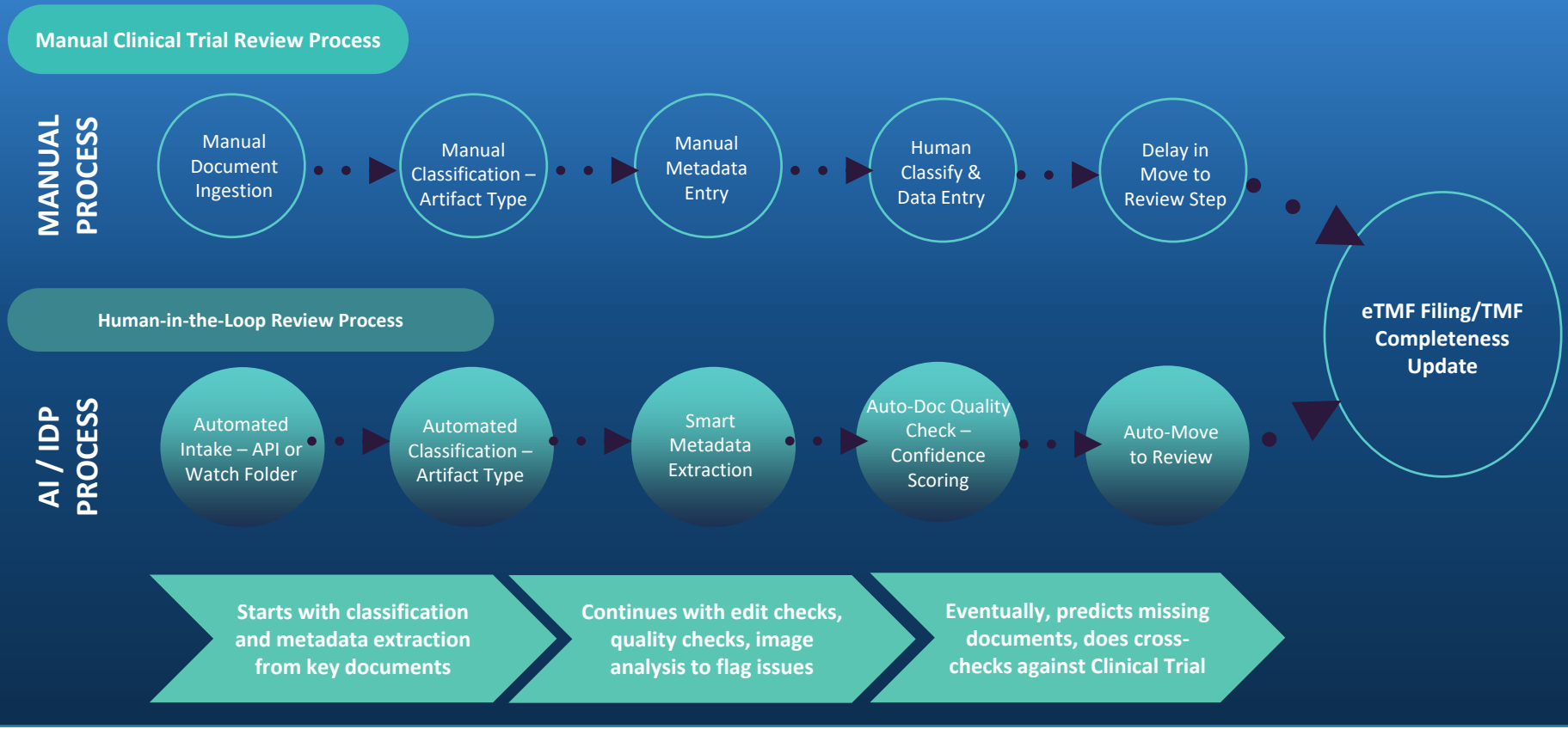
4.1.1.1 Study Object (ResearchStudy) JSON Example

```
{
  "resourceType": "ResearchStudy",
  "id": "study-001",
  "identifier": [
    {
      "use": "official",
      "system": "https://clinicaltrials.gov",
      "value": "NCT01234567",
      "type": {
        "coding": [
          {
            "system": "http://terminology.hl7.org
            "code": "CT",
            "display": "Clinical Trial Number"
          }
        ]
      },
      "assigner": {
        "display": "ClinicalTrials.gov"
      }
    }
  ],
  "title": "A Phase III Randomized Study of Drug
  "status": "active",
  "description": "This study aims to evaluate the
hypertension.",
  "primaryPurposeType": {
    "coding": [
      {
        "system": "http://terminology.hl7.org/Co
        "code": "treatment",
```

4.2.1.1 Trial Document Object (DocumentReference) JSON Example

```
{
  "resourceType": "DocumentReference",
  "id": "doc-001",
  "identifier": [
    {
      "use": "official",
      "system": "http://example.org/documents",
      "value": "ICF-123456"
    }
  ],
  "status": "current",
  "type": {
    "coding": [
      {
        "system": "http://loinc.org",
        "code": "34108-1",
        "display": "Informed Consent Form"
      }
    ]
  },
  "description": "Informed Consent Form for Clinical Trial",
  "extension": [
    {
      "url": "http://example.org/fhir/StructureDefinition/tmf-level",
      "valueString": "Site"
    },
    {
      "url": "http://example.org/fhir/StructureDefinition/tmf-zone",
      "valueString": "Central Trial Documents"
    },
```

AI “Hyperautomation” in the TMF Workflow



Total Libraries	Total Prompts	Avg Response Time	Token Usage	System is Healthy
24	156	24s	1.4M	Tue, Sep 30, 2025
+2 this month	+2 this month	-0.3s today	+12% this month	All systems are running normally.

Doc Accuracy	Doc Issues	Docs with PHI	Doc Throughput
86.5%	120	1.2K	42/sec
+2 this month	20% of all Docs	40% of all Docs	-0.3s today

Doc Accuracy of Each Type

DOCUMENT TYPE	TOTAL DOC	ACCURACY	STATUS	REQUIRED ACTION
Acceptance Of Investigator's Brochure	15,000	98%	Good	-
CAP Certificate	5,000	98%	Good	-
CLIA Certificate	12,000	100%	Good	-
Expedited Safety Report	300	99%	Good	-
Form FDA 1572	1,200,000	92%	Good	-
Informed Consent Form	1,240,000	76%	Critical	Investigate Results
Monitoring Visit Report	7,000,000	100%	Good	-
Protocol Signature Page	1,200,000	81%	Critical	Investigate Results

Status Notifications

- **yulian** Created new prompt library **eTMF Library** — 32m ago
- **maya** Updated user **dmytro** — 45m ago
- **cirill** Created new prompt **Meeting Minutes Template** — 1h ago
- **jordan** Created new prompt **Form FDA 1527** — 2h ago
- **cirill** Created new prompt **Patient Consent Form** — 3h ago
- **maya** Updated prompt **Clinical Study Protocol** — 4h ago
- **jordan** Created new prompt library **Financial Reports** — 6h ago
- **dmytro** Updated prompt **Protocol** — 1d ago

Benefits of a Pure Metadata Approach

Reduced data entry	Across all integrated systems. Improved inspection, searching.
De-risk integrations	As above, data is fungible across ecosystem. Integrations become low risk or unnecessary.
More, higher quality metadata	Quality improves with automation. Extraction often limited by human effort.
AI foundations	Deterministic automation (completeness checks). Probabilistic AI (insights, risk detection).
Digital → intelligence	Document extraction still requires HITL support automation. AI/LLM requires data for intelligence.

Lessons Learned from Past Initiatives

ODM(Operational Data Model)

- Highly prescriptive technology approach (XML)
- Completely flexible structure
- Designed for data, not documents
- Focused on metadata and data model
- Focused use cases on subject data (not comprehensive)
- High machine readability (can be parsed easily by any software)
- Allows expansion and extensions
- Poor human readability
- Industry adoption is optional and strong

ICH eCTD (Electronic Common Technical Document)

- Highly prescriptive technology approach (XML)
- Highly prescriptive index structure
- Designed for submissions and documents, not data
- Comprehensive use cases to capture all submission types (except med device)
- Poor machine readability (requires specialized software)
- Limited flexibility, not expandable
- Poor human readability
- Industry adoption is weak but enforced

TMF Digitalization Roadmap

2024–2025
- Public roadmap published
- Vendor panel started
- Interoperability discussions

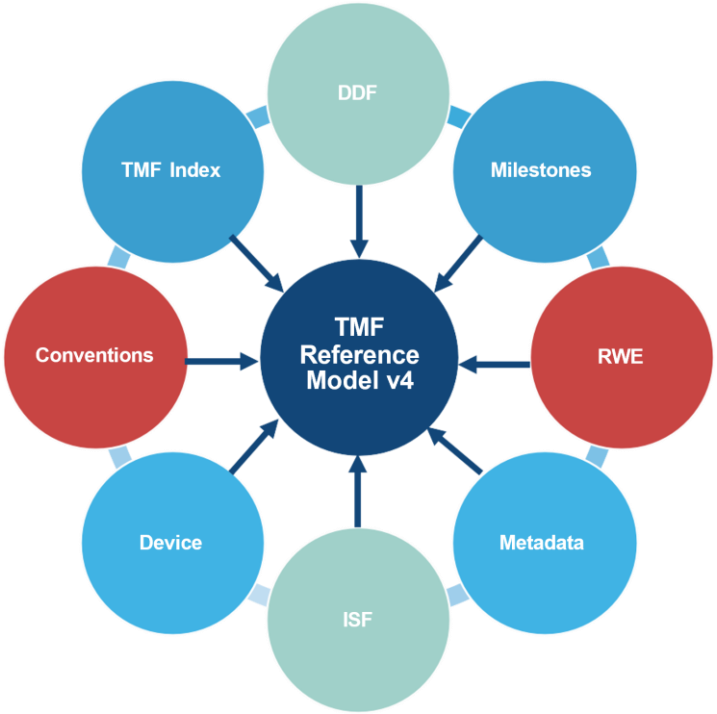
2026
- Public review/comment period
- Controlled terminology finalized

2027
- CDISC TMF RM4 released
- USDN 1.0

(Possible) Future
- Merge with DDF/USDN
- Data-first, metadata-centric TMF

Actions and Next Steps

- Sponsors and CROs should consider their entire clinical ecosystem, going beyond their four walls
- Vendors should consider data standards as a baseline requirement for their systems
- Can focus to solve iterative, pragmatic use cases that solve real operational business challenges
- Should look to view these standards as the pragmatic baseline for system interoperability
- Eventually, should look to extend TMF into DDF and align with USDN



Thank You!