REVIEW

Data Before Models: Building a Surgical Data Layer

Subhas Gupta, MD, CM, PhD, FRCSC, FACS

Department of Plastic Surgery, Loma Linda University; The Plastic Surgery AI Group

Published November 2025 · DOI: 10.xxxx/jsi.2025.0002

Abstract. Every AI system in surgery, from imaging models to workflow agents, depends on structured, high-quality data. Yet most surgical environments operate with scattered photographs, fragmented electronic records, and free-text notes. This article defines how to build a Surgical Data Layer: a unified, auditable foundation that allows models to learn safely, perform accurately, and improve continuously.

Keywords: surgical data layer, data governance, de-identification, machine learning, data quality

1. Why Data Quality Determines Clinical Value

AI is not magic; it is mathematics built on data. Inconsistent inputs lead to unreliable outputs. Across all surgical specialties, the most successful AI deployments share three traits: standardized data capture across photos, imaging, notes, and sensors; traceability of when, how, and by whom data was created; and feedback loops that link outcomes back into the dataset.

Plastic surgery already models this well, with consistent photography, anatomical labeling, and detailed complication logs that naturally support accurate model training.

2. What Is a Surgical Data Layer?

A Surgical Data Layer is not a new database. It is an architecture that connects, cleans, and labels existing systems so that AI can work reliably, organized across four levels:

Level	Description	Example
Raw Data	Source files and text	DICOMs, clinic photos, operative notes
Structured Layer	Labeled and standardized	View labels, anatomical tags, FHIR text
Feature Layer	Model-ready inputs	Tissue type, laterality, operative time
Application Layer	End-user outputs	Dashboards, predictions, documentation

This architecture transforms routine clinical data into learning data for ongoing improvement.

3. Practical Steps for Building a Surgical Data Layer

- Map all data sources: clinic photos, OR video, anesthesia logs, EMR exports, PROMs, wearable data.
- Standardize formats: consistent file types, resolutions, naming conventions, and metadata.
- De-identify at ingestion: automate removal of PHI while preserving clinical meaning.
- Tag meaningfully: include laterality, anatomy, view, date, and procedural metadata.
- Create version control: track how data changes over time; never overwrite clinical history.
- Integrate outcomes: link complications, revisions, and patient-reported outcomes back to the dataset.

Plastic surgery groups often start with standardized photography (frontal, lateral, oblique) then expand to structured operative documentation.

4. Governance and Ethics

A Surgical Data Layer must be governed with the same care as any clinical system: HIPAA and GDPR compliance by design, access controls for labeling and export, audit trails for every modification, and equity checks to ensure inclusive representation of patient populations. Academic departments and large practices should appoint data stewards, clinicians responsible for labeling accuracy and dataset integrity.

5. Return on Investment

A well-structured data layer reduces cost, improves safety, and accelerates innovation through fewer failed model deployments, reusable datasets for research and quality improvement, faster regulatory validation, improved workflow efficiency, and opportunities for cross-department collaboration. Groups with mature data layers are already deploying tools for scar prediction, flap viability assessment, and operative planning, built internally without relying on external vendors.

Key Takeaways

- A Surgical Data Layer turns clinical information into a strategic asset.
- Clean, traceable, well-labeled data is the foundation of trustworthy surgical AI.
- Plastic surgery's imaging discipline can guide other surgical fields.