Evolution of the Mount Sinai Data Warehouse (MSDW)

Scientific Computing
Icahn School of Medicine at Mount Sinai

November 2021

Icahn
School of
Medicine at
Mount
Sinai

Motivation for MSDW Evolution

1. Legacy MSDW not aligned with Mt Sinai's **Epic strategy**
   - Ever more data already in Epic Caboodle data warehouse
   - Hard-to-query bespoke data model no longer justified

2. **Technological limitations** of legacy MSDW
   - Bottlenecks in loading data as volumes grew
   - Opportunity to switch to cheaper database & software

3. Need to **streamline data sharing**
   - Many research networks & programs now using OMOP Common Data Model
   - Leverage tools & resources of OHDSI community

Legacy MSDW in 2010 versus 2020

	MSDW Circa 2010	MSDW Circa 2020
Data Sources	Quickly acquire data from myriad systems by copying **HL7 message** transmissions	Ever more of Mt Sinai's data is/will be in **Epic**
Common Data Models	**Limited data sharing** in research networks (OHDSI not founded until 2013)	Need to support **extensive data sharing** with CTSA, Insight CRN, TriNetX, Sema4 using common data models like **OMOP**
Data Model	EAV-style **single fact table**, similar to i2b2, for easy data ingestion with few resources	Very **difficult** to reconstruct clinical events & transactions, slows data delivery, obscures data lineage
Data Pipeline	Designed to **overcome bottlenecks**	Bottlenecks have **re-emerged**, daily loads are becoming infeasible
Technology Stack	Oracle technology stack aligned with Mt Sinai's purchase of **Oracle Healthcare Analytics (OHA)** platform	Oracle & SAP tools are expensive, no longer aligned with Mt Sinai's **Epic strategy**

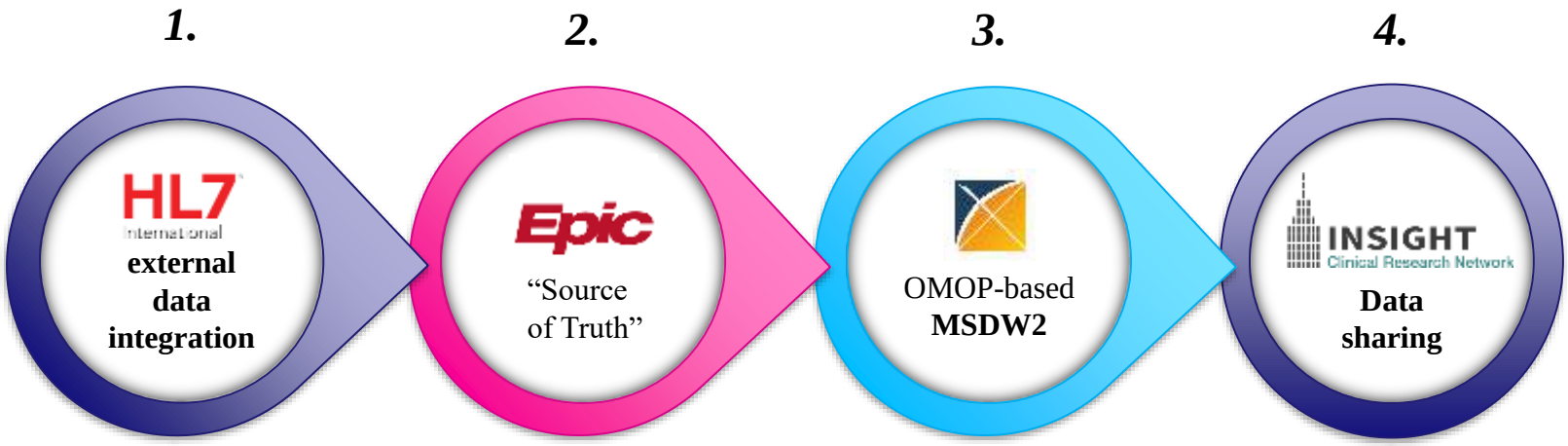
Goals for MSDW2

1. **Improved researcher access & data quality**
   - Self-service query capabilities from OMOP and TriNetX
   - Better integration with other data sets on Minerva
   - Transparency of data lineage & data processing
   - Higher efficiency, quicker data request turnaround

2. **Faster data refreshes**
   - Enable daily incremental updates

3. **Sustainable data infrastructure**
   - Minimize cost of integration, maintenance, data governance

Epic & OMOP-centric Solution for MSDW2



- **Most economical** by leveraging Epic product & IT investment
- **Best solution** to address prior shortcomings
- **Best foundation** on which to build new capabilities

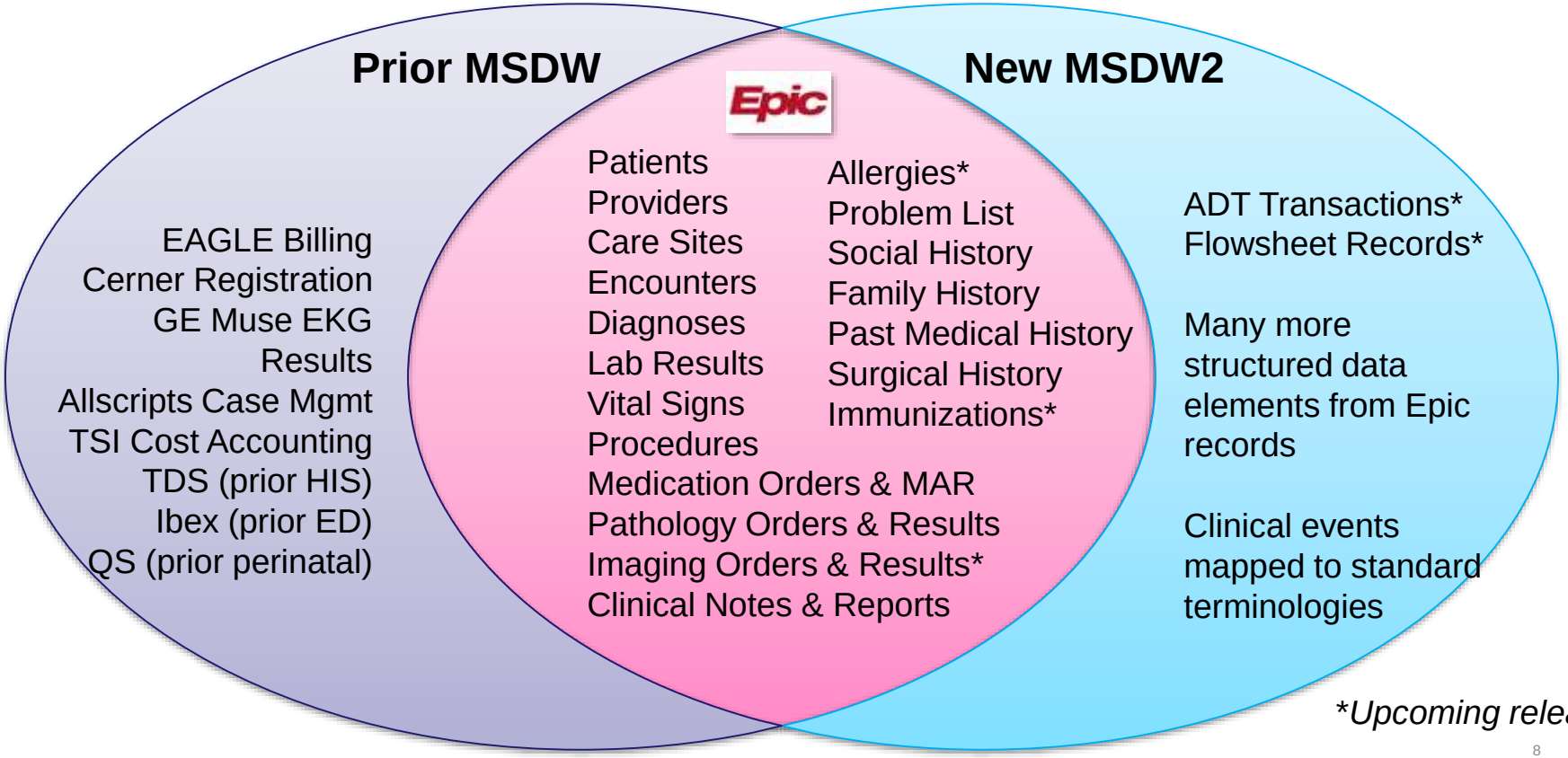
How We're Evolving MSDW

	Legacy MSDW	New MSDW2
Data Sources	Primarily **HL7 messages**; Epic added later	**Epic** as primary data source
Data Model	**Bespoke model** for easy data ingestion, but difficult to query	**OMOP Common Data Model** for standardized structure & content, enables data sharing
Data Pipeline	Original design **no longer supports** daily loads	Restore **daily load** capability, with **full change history**
Technology Stack	Originally aligned with Mt Sinai's purchase of **Oracle Healthcare Analytics (OHA)** platform	Align with **Epic Caboodle** and **Minerva HPC**
Costs	Licenses and labor costs are **unsustainable**	Leverage Mount Sinai's **Epic investment**

Benefits to Mount Sinai

MSDW Evolution	Benefits
Epic as Primary Data Source	• Higher data quality • Transparent data lineage • Access to more data elements
OMOP Common Data Model	• The *de facto* standard data model for research • Standardized structure & content • Vibrant research community & tools • Easier data sharing with research networks
Restore Daily Refreshes	• Faster access to data • Full history of all changes
Change in Technologies	• Lower cost to Mount Sinai • Enable faster development of new features

Gap Analysis: Prior MSDW vs New MSDW2



**Upcoming release*