

Agile Mastering: A New Way of Thinking about MDM

Michael Doane, Solutions Director, MarkLogic



MANY DEFINITIONS

What is Master Data Management?

"MDM is the consistent and uniform set of identifiers and extended attributes that describe the core entities of the enterprise and are used across multiple business processes."

— Gartner

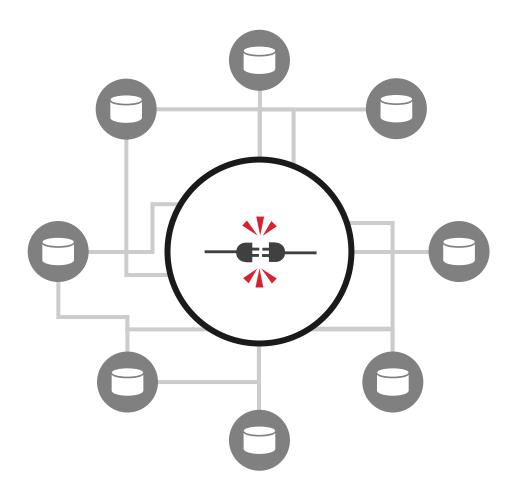
"The set of disciplines and methods to ensure the currency, meaning, quality, and deployment of a company's reference data within and across subject areas."

Baseline Consulting

"A set of disciplines, processes and technologies, for ensuring the accuracy, completeness, timeliness and consistency of multiple domains of enterprise data – across applications, systems and databases, and across multiple business processes, functional areas, organizations, geographies and channels."

— Dan Power, Hub Designs

■ MarkLogic •



WHY MDM IS NEEDED

Getting a Common View Across Data Silos

- Patients
- Customers
- Products
- Pharmaceuticals

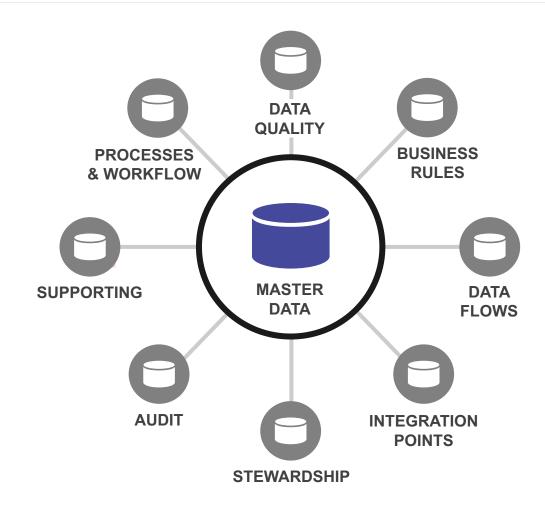
MarkLogic MarkLogic

Mastering Data Makes Sense

... But ...



MANY COMPONENTS MDM is Complex



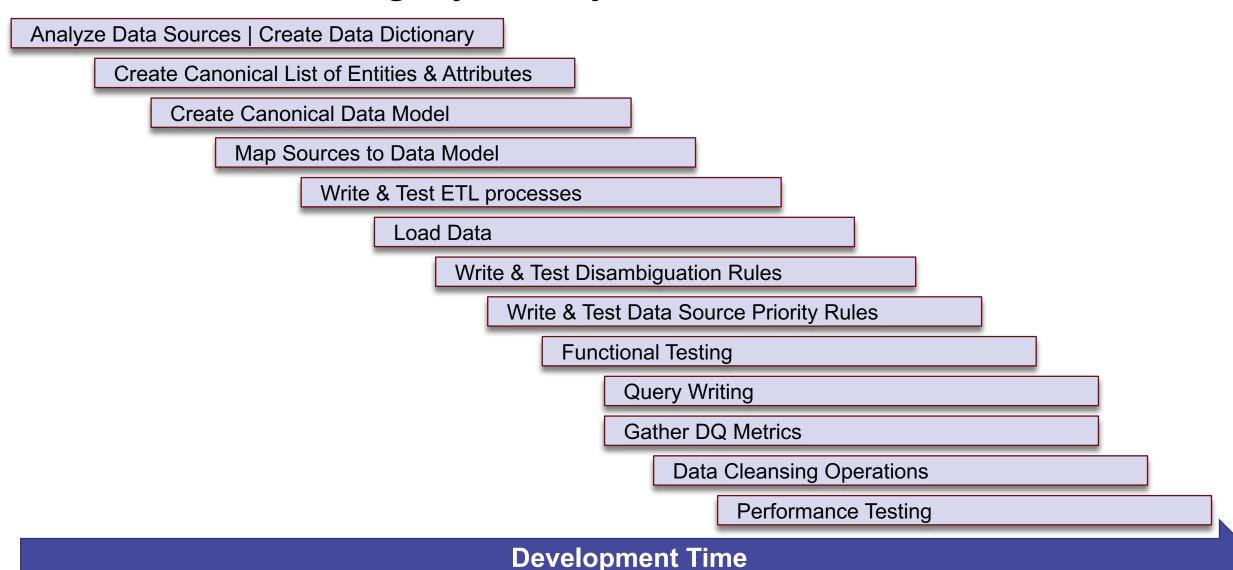


Traditional MDM Faces Strong Headwinds

- Lengthy modeling & ETL design processes = slow progress, dwindling interest, brittle to adapt to changes in goals and data
- Some successes in personnel and custom siloed solutions
- Usually owned by IT and disconnected from the impact to the enterprise's profit and loss
- Chasing a truth-based model with a rigid golden definition vs. a trust-based model with a golden portion and flexibility to capture and keep all data

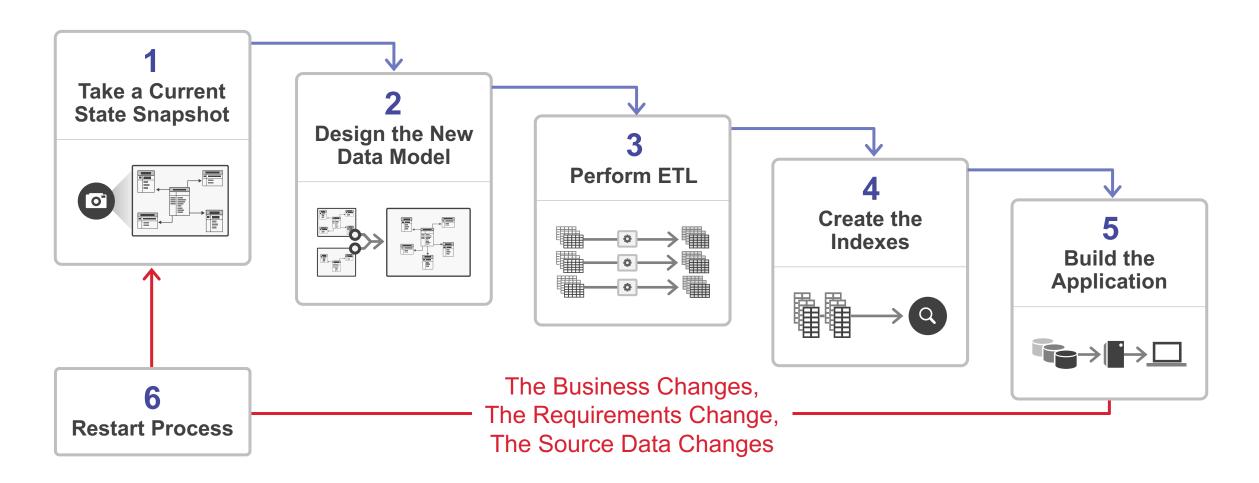


Traditional Lengthy Lifecycle of RDBMS-based MDM



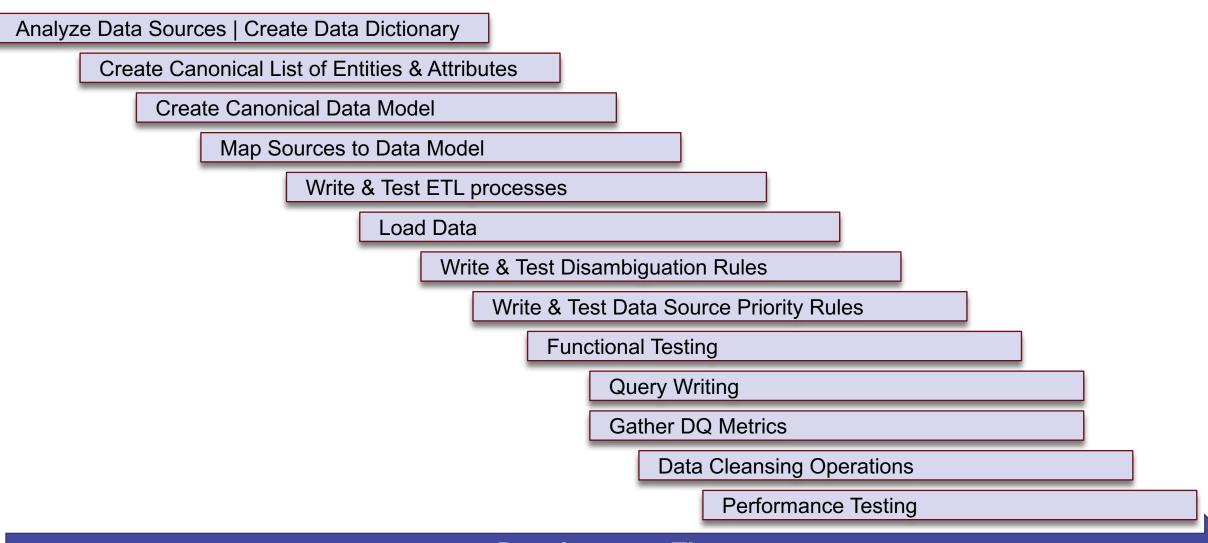


The Problem With the Relational Approach





Repeat This Again?

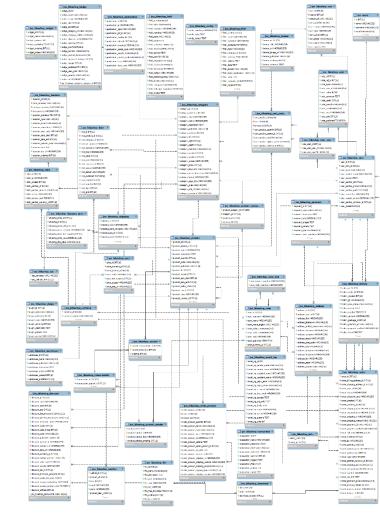


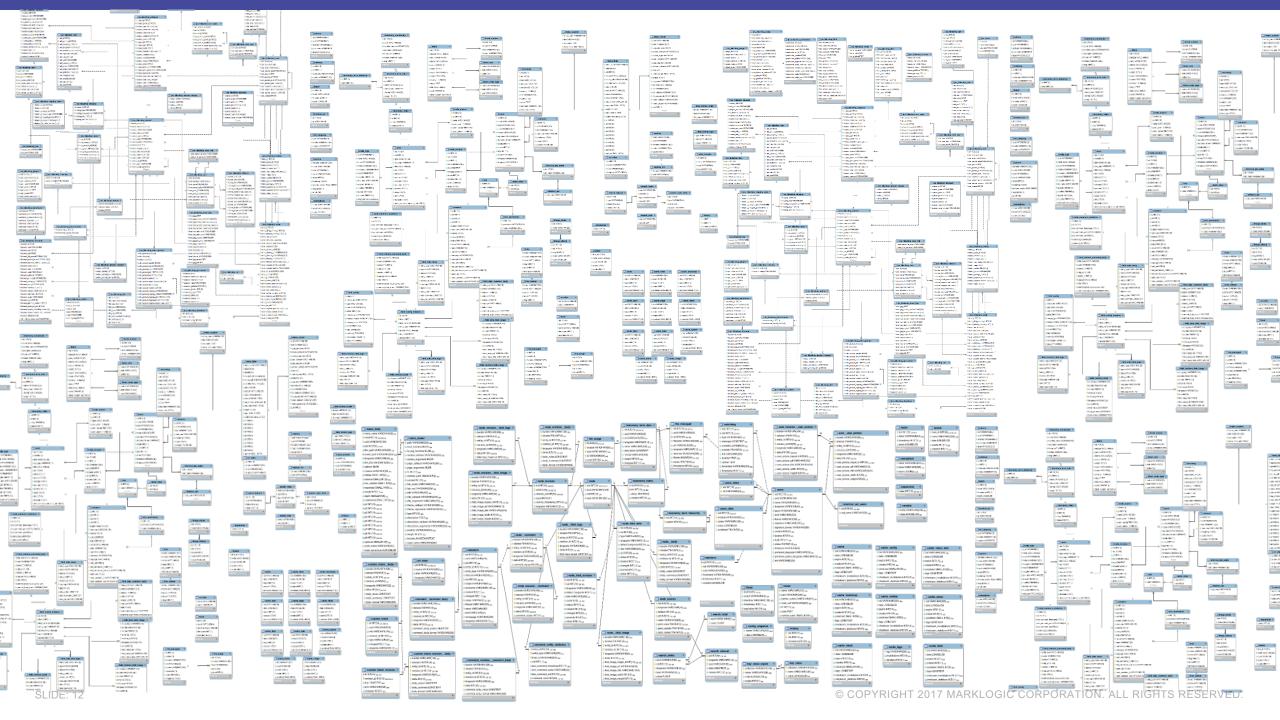
Development Time

MarkLogic MarkLogic

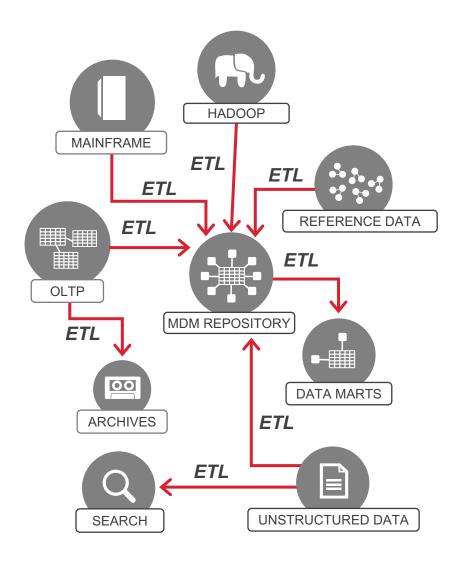
STILL UNSURE WHY IT'S SO HARD?







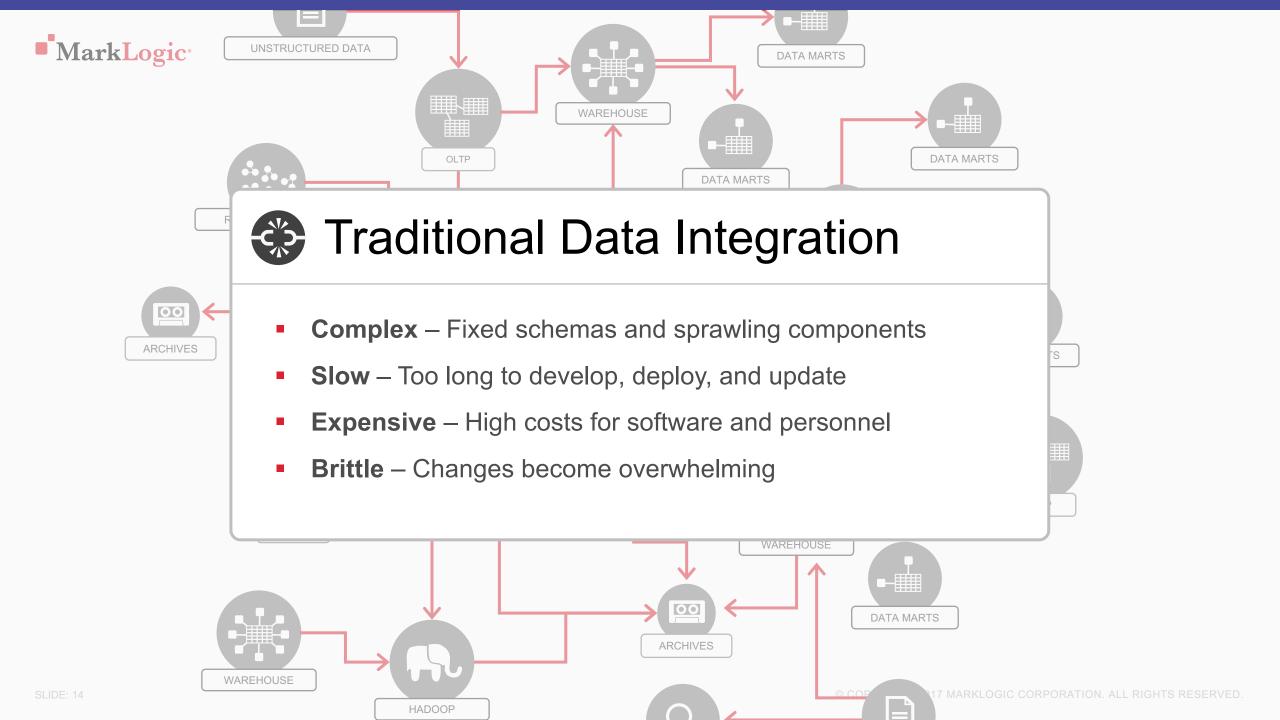
■ MarkLogic •



UNFORTUNATE REALITY

Traditional Data Integration Is Complex and Fractured

- RDBMS for highly structured data
- Specialized databases for other data types
- ETL and integration software to connect silos
- ...and, what about hierarchical data?
- ...and, what about unstructured content?
- ...maybe a data lake will help?





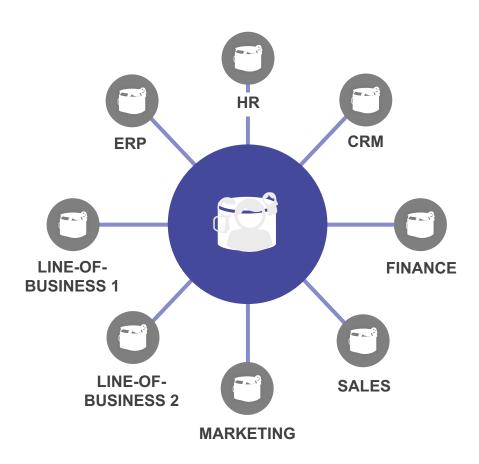


A NEW APPROACH

Re-thinking Master Data Management

- Achieve progress incrementally
- Tie to business drivers and events
- Reduce duplicate data and lengthy ETL processes
- Adjust to change

■ MarkLogic •



FIX THE ARCHITECTURE

Operational Data Hub

- 360 views of important entities
- Direct integration with transactional applications
- Real-time
- Handle volume, variety, velocity (like a data lake)

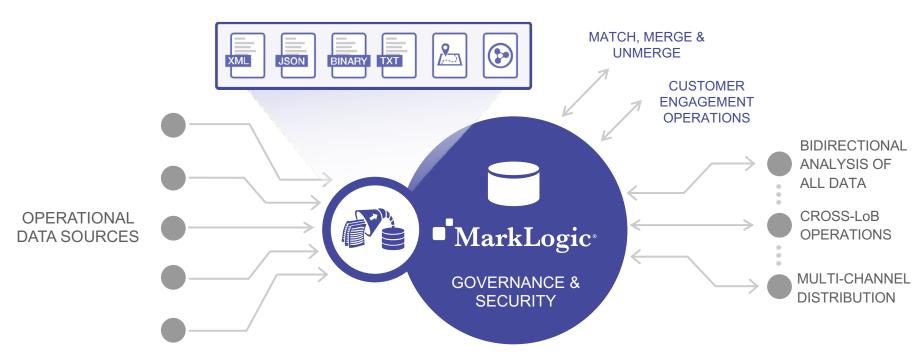


A NEW APPROACH TO MDM

Agile Mastering

ODH-Integrated: Supports the trend towards multi-domain MDM

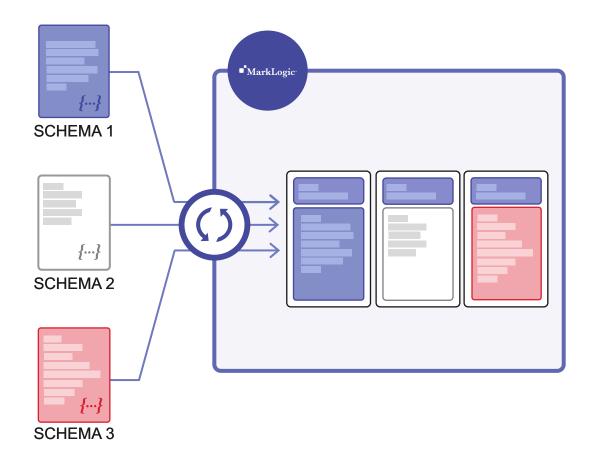
Go from 'collecting' to 'connecting' data for all your domains



OPERATIONAL DATA HUB CORE

NEW SOURCES EASILY ADDED HARMONIZE IN PLACE

• MarkLogic



A NEW APPROACH

Agile Mastering

Schema-agnostic: Load "as-is", minimize ETL, incrementally deliver results early on, maintain buy-in



CUSTOMER ENGAGEMENT OPERATIONS



A NEW APPROACH

Agile Mastering

Contextual: Improved data quality through minimizing data duplication at point of engagement



```
<?xml version="1.0" encoding="UTF-8"?>
 <?xml version="1.0" encoding="UTF-8"?>
   <es:envelope xmlns:es="http://marklogic.com/entity-services">
     <document xmlns="http://www.w3.org/ns/prov#">
       <activity xmlns:prov="http://prov#"
       prov:id="/auditing#merge/e4c664e1-ffe2-457d-b73a-46fb7a2d2625.xml-
       15280186555938034808">
         <type xsi:type="xsd:string">merge</type>a
         <label>merge by admin</label>
       </activity>
       <entity xmlns:prov="http://prov#" prov:id="/auditing#/Person/208.xml">
         <type xsi:type="xsd:string">contributing record for merge</type>
         <label>/CSV FILE/Person/208.xml</label>
       </entity>
       <entity xmlns:prov="http://prov#" prov:id="/auditing#/Person/199.xml">
         <type xsi:type="xsd:string">contributing record for merge</type>
         <label>/CSV_FILE/Person/199.xml</label>
       <entity xmlns:prov="http://prov#" prov:id="/auditing#//merged/2d2625.xml">
         <type xsi:type="xsd:string">result of record merge</type>
         <label>/com.marklogic.agile-mastering/merged/2-457d-b73a-46fb7a2d2625.xml
         </label>
                                                          SIDECAR
       </entity>
       <wasDerivedFrom>
         <generatedEntity xmlns:prov="http://prov#"</pre>
         prov:ref="/auditing#/merged/46fb7a2d2625.xml" />
         <usedEntity xmlns:prov="http://prov#" prov:ref="/auditing#/Person/208.xml"/>
         <activity xmlns:prov="http://prov#"</pre>
         prov:ref="/auditing#merge/2-457d-b73a-46fb7a2d2625.xml-15280186555938034808>
       </wasDerivedFrom>
       <agent xmlns:prov="http://prov#"</pre>
       prov:id="http://marklogic.com/agile-mastering/auditing#user-admin"> com </agent>
       <wasAttributedTo> co </wasAttributedTo>
       <wasGeneratedBy>  
       <entity xmlns:prov="http://prov#"</pre>
       prov:id="/auditing#/CSV_FILE/Person/199.xmlPersonNamebcc184075b"> cm </entity>
       <entity xmlns:prov="http://prov#"</pre>
       prov:id="/auditing#/CSV_FILE/Person/199.xmlidcf83e135"> cm </entity>
       /triplestore/5b8e174236b572b3.xml
       <sem:triples xmlns:sem="http://marklogic.com/semantics">
       <sem:triple><sem:subject>http://agile-mastering/auditing#attribution-admin-
       917c1c40-d2b8-404692ec95d33ae1bbf8</sem:subject>
       <sem:predicate>http://www.w3.org/1999/02/22-rdf-syntax-ns#type</sem:predicate>
       <sem:object>http://www.w3.org/ns/prov#Attribution</sem:object>
       </sem:triple>
         <sem:triple> co </sem:triple>
         <sem:triple> co </sem:triple>
       </sem:triples>
```

A NEW APPROACH

Agile Mastering

All the Data: Master a subset and keep all source data in original formats; reverse changes or unmerge at any time



Agile Mastering Pre-Merge Source Docs

```
<?xml version="1.0" encoding="UTF-8"?>
<envelope xmlns="entity-services" xmlns:agile-mastering="/agile-mastering">
    <agile-mastering:id>a668214b-210f-4b66-8ead-0d11081f7005</agile-mastering:id>
    <agile-mastering:sources>
     <agile-mastering:source>
       <agile-mastering:name>CSV_FILE</agile-mastering:name>
       <agile-mastering:import-id>mdm-import-1clae2ce-79d4-470f-a2e5-a5</agile-mastering:import-id>
       <agile-mastering:user>admin</agile-mastering:user>
       <agile-mastering:dateTime>2017-03-27T22:36:39.624185Z</agile-mastering:dateTime>
     </agile-mastering:source>
    </agile-mastering:sources>
  </headers>
                                          SOURCE #1
  <triples/>
  <instance>
    <MDM xmlns="">
     <Person>
       <PersonName>
             <PersonGivenName>JILL</PersonGivenName>
             <PersonSurName>DEMODEC</PersonSurName>
           </PersonName>
          <id>199</id>
           <Address>
             <LocationState>MD</LocationState>
             <AddressPrivateMailboxText>9999</AddressPrivateMailboxText>
             <AddressSecondaryUnitText>VIRGINIA</AddressSecondaryUnitText>
             <LocationPostalCode>215020000</LocationPostalCode>
             <LocationCityName>CUMBERLAND</LocationCityName>
          <PersonSSNIdentification>
             <IdentificationID>7377777</IdentificationID>
           </PersonSSNIdentification>
       </Person>
    </MDM>
  </instance>
 <attachments>
    <person xmlns="">
     <LIVING ARRANGEMENT TYPE CODE>AH</LIVING ARRANGEMENT TYPE CODE>
     <CL_FIRST_NAME>JILL</CL_FIRST_NAME>
     <STREET TYPE CODE>AVE</STREET TYPE CODE>
     <AU_RR_PRD_START_DT>201606</AU_RR_PRD_START_DT>
     <CLIENT_ID>199</CLIENT_ID>
     <DOB>19900101</DOB>
     <ZIPCODE RES>215020000</ZIPCODE RES>
     <UI FREQ CD>-</UI FREQ CD>
     <STREET 2 ADDRESS RES/>
     <DISABILITY_TYPE_CODE>-</DISABILITY_TYPE_CODE>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<envelope xmlns="/entity-services" xmlns:agile-mastering="/agile-mastering">
   <agile-mastering:id>3d39141a-ad6e-44b3-8884-8b6f685727e6</agile-mastering:id>
   <agile-mastering:sources>
     <agile-mastering:source>
       <agile-mastering:name>CSV FILE</agile-mastering:name>
       <agile-mastering:import-id>mdm-import-1c1ae2ce-79d4-470f-a2e5-a5</agile-mastering:import-id>
       <agile-mastering:user>admin</agile-mastering:user>
       <agile-mastering:dateTime>2017-03-27T22:36:39.624185Z</agile-mastering:dateTime>
     </agile-mastering:source>
   </agile-mastering:sources>
                                         SOURCE #2
 </headers>
 <triples/>
 <instance>
   <MDM xmlns="">
     <Person>
             <PersonGivenName>JILLIAN</PersonGivenName>
             <PersonSurName>DEMODEC</PersonSurName>
           </PersonName>
         <id>208</id>
           <Address>
             <LocationState>MD</LocationState>
             <AddressPrivateMailboxText>9999</AddressPrivateMailboxText>
             <AddressSecondaryUnitText>VIRGINIA</AddressSecondaryUnitText>
             <LocationPostalCode>215020000
             <LocationCityName>CUMBERLAND</LocationCityName>
           </Address>
         <PersonSSNIdentification>
             <IdentificationID>7377777</IdentificationID>
           </PersonSSNIdentification>
       </Person>
   </MDM>
 </instance>
 <attachments>
   <person xmlns="">
     <LIVING_ARRANGEMENT_TYPE_CODE>AH</LIVING_ARRANGEMENT_TYPE_CODE>
     <CL FIRST NAME>JILLIAN</CL FIRST NAME>
     <STREET TYPE CODE>AVE</STREET TYPE CODE>
     <AU_RR_PRD_START_DT>201606</AU_RR_PRD_START_DT>
     <CLIENT ID>208</CLIENT ID>
     <DOB>19900101</DOB>
     <ZIPCODE_RES>215020000</ZIPCODE_RES>
     <UI FREQ CD>-</UI FREQ CD>
     <STREET 2 ADDRESS RES>
     </STREET 2 ADDRESS RES>
```



Agile Mastering Merged – All The Data - Sidecar Doc

```
<es:envelope xmlns:es="http://marklogic.com/entity-services">
 <es:headers>
    <agile-mastering:id xmlns:agile-mastering="/agile-mastering">
            27466526-12b6-4af1-8e76-3ad247d6c08e
    </agile-mastering:id>
    <agile-mastering:merges xmlns:agile-mastering="http://marklogic.com/agile-mastering">
     <agile-mastering:document-uri>/CSV FILE/Person/208.xml</agile-mastering:document-uri>
     <agile-mastering:document-uri>/CSV FILE/Person/199.xml</agile-mastering:document-uri>
    </agile-mastering:merges>
    <agile-mastering:sources xmlns:agile-mastering="http://marklogic.com/agile-mastering">
     <agile-mastering:source>
       <agile-mastering:name>CSV FILE</agile-mastering:name>
       <agile-mastering:import-id>mdm-import-a9251bcb-a9ee-45f5-88a8-fc</agile-mastering:import-id>
       <agile-mastering:user>admin</agile-mastering:user>
       <agile-mastering:dateTime>2017-04-06T18:22:14.158686Z</agile-mastering:dateTime>
     </agile-mastering:source>
     <agile-mastering:source>
       <agile-mastering:name>CSV FILE</agile-mastering:name>
       <agile-mastering:import-id>mdm-import-a9251bcb-a9ee-45f5-88a8-fc</agile-mastering:import-id>
       <agile-mastering:user>admin</agile-mastering:user>
       <agile-mastering:dateTime>2017-04-06T18:22:14.158686Z</agile-mastering:dateTime>
     </agile-mastering:source>
                                              MERGED
    </agile-mastering:sources>
  </es:headers>
 <es:instance>
    <MDM>
     <Person>
             <LocationState>MD</LocationState>
             <AddressPrivateMailboxText>9999</AddressPrivateMailboxText>
             <AddressSecondaryUnitText>VIRGINIA</AddressSecondaryUnitText>
             <LocationPostalCode>215020000</LocationPostalCode>
             <LocationCityName>CUMBERLAND/LocationCityName>
           </Address>
           <PersonName>
             <PersonGivenName>JILLIAN</PersonGivenName>
             <PersonSurName>DEMODEC</PersonSurName>
           <PersonSSNIdentification>
             <IdentificationID>7377777</IdentificationID>
           </PersonSSNIdentification>
         <id datatype="array">208</id>
         <id datatype="array">199</id>
       </Person>
    </MDM>
 </es:instance>
</es:envelope>
```

```
<document xmlns="http://www.w3.org/ns/prov#">
 <activity xmlns:prov="http://prov#"
 prov:id="/auditing#merge/e4c664e1-ffe2-457d-b73a-46fb7a2d2625.xml-
 15280186555938034808">
   <type xsi:type="xsd:string">merge</type>a
   <label>merge by admin</label>
 <entity xmlns:prov="http://prov#" prov:id="/auditing#/Person/208.xml">
   <type xsi:type="xsd:string">contributing record for merge</type>
   <label>/CSV FILE/Person/208.xml</label>
 <entity xmlns:prov="http://prov#" prov:id="/auditing#/Person/199.xml">
   <tvpe xsi:tvpe="xsd:string">contributing record for merge</tvpe>
   <label>/CSV FILE/Person/199.xml</label>
 </entity>
 <entity xmlns:prov="http://prov#" prov:id="/auditing#//merged/2d2625.xml">
   <type xsi:type="xsd:string">result of record merge</type>
   <label>/com.marklogic.agile-mastering/merged/2-457d-b73a-46fb7a2d2625.xml
   </label>
                                                 SIDECAR
 </entity>
 <wasDerivedFrom>
   <generatedEntity xmlns:prov="http://prov#"</pre>
   prov:ref="/auditing#/merged/46fb7a2d2625.xml" />
   <usedEntity xmlns:prov="http://prov#" prov:ref="/auditing#/Person/208.xml"/>
   <activity xmlns:prov="http://prov#"
   prov:ref="/auditing#merge/2-457d-b73a-46fb7a2d2625.xml-15280186555938034808>
 </wasDerivedFrom>
 <wasDerivedFrom> come </wasDerivedFrom>
 <agent xmlns:prov="http://prov#"</pre>
 <wasAttributedTo> co </wasAttributedTo>
 <wasGeneratedBy> colored
 <entity xmlns:prov="http://prov#"</pre>
 <wasDerivedFrom> come </wasDerivedFrom>
 <entity xmlns:prov="http://prov#"</pre>
 prov:id="/auditing#/CSV FILE/Person/199.xmlidcf83e135"> cm </entity>
 /triplestore/5b8e174236b572b3.xml
 <sem:triples xmlns:sem="http://marklogic.com/semantics">
 <sem:triple><sem:subject>http://agile-mastering/auditing#attribution-admin-
 917c1c40-d2b8-404692ec95d33ae1bbf8</sem:subject>
 <sem:predicate>http://www.w3.org/1999/02/22-rdf-svntax-ns#type</sem:predicate>
 <sem:object>http://www.w3.org/ns/prov#Attribution</sem:object>
 </sem:triple>
   <sem:triple> co </sem:triple>
   <sem:triple> co </sem:triple>
 </sem:triples>
```

■ MarkLogic •

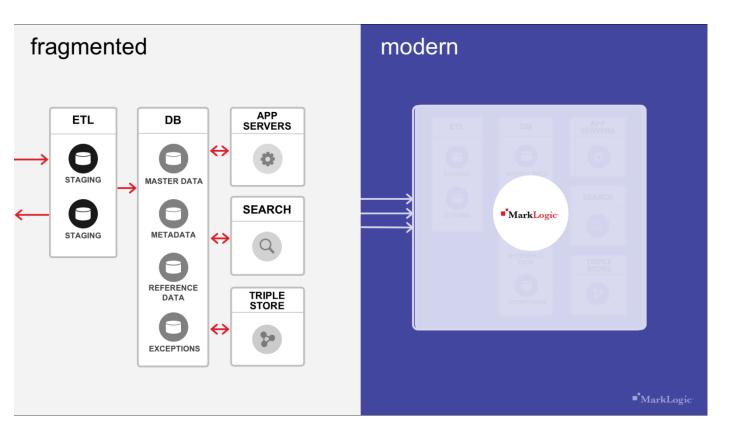
```
{ "metadata" : {
                                              { "metadata" : {
     "Source": "Finance",
                                                    "Source": "POS",
     "Date": "2016-04-17",
                                                    "Date": "2016-04-17",
    "Lineage": "v01 transform" },
                                                    "Lineage": "v01 transform" },
  "canonical": { "Zip": [94111] },
                                                 "canonical": { "Zip": [94111,94070] },
  "source": { "ID": 1001,
                                                 "source": { "Customer ID": 2001,
     "Fname": "Paul",
                                                    "Given Name": "Karen",
     "Lname": "Jackson",
                                                    "Family Name": "Bender",
     "Phone": "415-555-1212 | 415-555-1234",
                                                    "Shipping Address": {
     "SSN": "123-45-6789",
                                                      "Street": "324 Some Road",
     "Addr": "123 Avenue Road",
                                                      "City": "San Francisco",
     "City": "San Francisco",
                                                      "State": "CA".
     "State": "CA",
                                                      "Postal": "94111",
     "Zip": 94111 } }
                                                      "Country": "USA" },
                                                    "Billing Address": {
{"metadata": {
                                                       "Street": "847 Another Ave".
  "Source": "TwitterDB",
  "Date": "2016-04-17",
                                                      "City": "San Carlos",
                                                      "State": "CA".
  "Lineage": "x23Trans",
                                                       "Postal": "94070",
  "canonical": "{Twitter": "@PaulJackson"}
                                                      "Country": "USA" },
{ "ID" : 9999 ,
                                                    "Phone": [
     "First": "Paul",
     "Last": "Jackson".
                                                      { "Type" : "Home" , "Number" : "415-555-6789" } ,
                                                      { "Type" : "Mobile" , "Number" : "415-555-6789" } ] } }
      "Handle": "@PaulJackson"
```

A NEW APPROACH

Agile Mastering

Metadata Unlimited: Maintain data provenance, change history, security on every data element



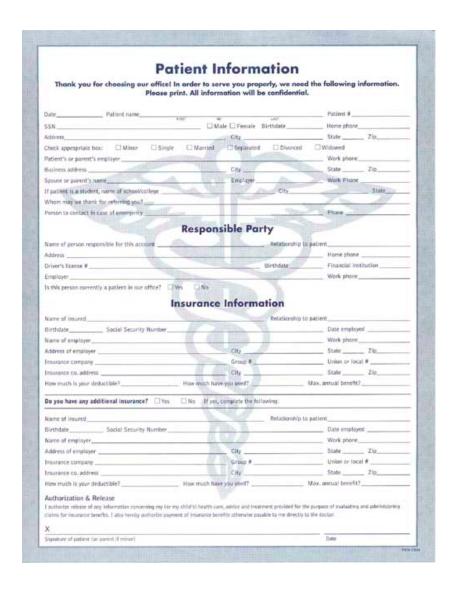


A NEW APPROACH

Agile Mastering

Simplified Architecture: Database, Search, Application Services, Security and more in one QA'd platform





A NEW APPROACH

Agile Mastering

Logical Data Model: Entities represent naturally as Documents, not shredded.

The data in your workflows is in documents!

Relate the documents logically and meaningfully with semantic triples

■ MarkLogic

```
<title>Continuity of Care Document</title>
<effectiveTime value="20140205020200" />
<confidentialityCode code="N" codeSystem="2.16.840.1.113883.5.25" />
<languageCode code="en-US"/>
<recordTarget><patientRole><!--Patient ID 000665834-01--><id</pre>
extension="000-66-5834" root="2.16.840.1.113883.4.1" />
<addr use="HP">
<streetAddressLine>1796 Country View Drive</streetAddressLine>
<city>Bloomingdale</city><state>NY</state><postalCode>12913</postalCode>
<telecom value="tel:+1-339-301-8947" use="HP"/>
<patient>
<name><given>Paula</given><given>A</given><family>Monroe</family>
</name>
<administrativeGenderCode code="F" displayName="Female"
codeSvstemName="HL7 AdministrativeGenderCodes"
codeSystem="2.16.840.1.113883.5.1" />
<birthTime value="19580318"/>
<maritalStatusCode code="D" displayName="Divorced"</pre>
codeSystem="2.16.840.1.113883.5.2" codeSystemName="MaritalStatusCode"/>
<raceCode displayName="White" code="2106-3"</pre>
codeSystem="2.16.840.1.113883.6.238" codeSystemName="CDC Race and
Ethnicity" />
<languageCommunication>
<languageCode code="en-US"/>
ferenceInd value="true"/>
</languageCommunication>
</patient>
<assignedAuthor><id extension="2658947110" root="2.16.840.1.113883.4.6" />
<addr use="WP">
<streetAddressLine>2204 Brandywine Street</streetAddressLine>
<city>Keene</city><state>NY</state><postalCode>12942</postalCode>
</addr>
<telecom value="tel:+1-320-784-2019" use="WP"/>
<assignedPerson>
<given>Geoffrey</given><family>Riggs</family><suffix>MD</suffix>
</name>
</assignedPerson>
<representedOrganization><id</pre>
root="2.16.840.1.113883.4.450" />
<name>ExactData County MedicalCenter
<telecom use="WP" nullFlavor="UNK" />
```

A NEW APPROACH

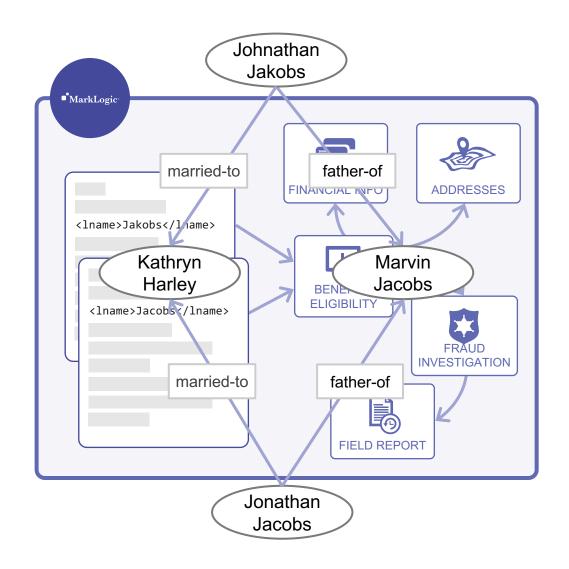
Agile Mastering

Logical Data Model: Entities represent naturally as Documents, not shredded.

The data in your workflows is in documents!

Relate the documents logically and meaningfully with semantic triples





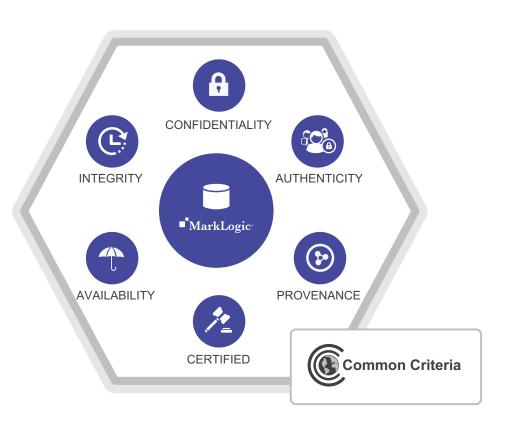
A NEW APPROACH TO MDM

Agile Mastering

Semantics: Capture rich relationships between entities to handle multiple domains

Improve matching



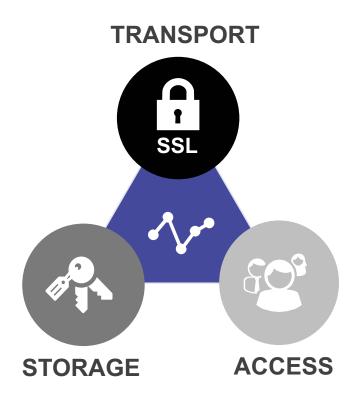


A NEW APPROACH

Agile Mastering

Secure: Support for operational apps data access control and data governance





A NEW APPROACH

Agile Mastering

Full Auditing: Capture who, what, when and why for all user and automated data changes and actions

"MarkLogic

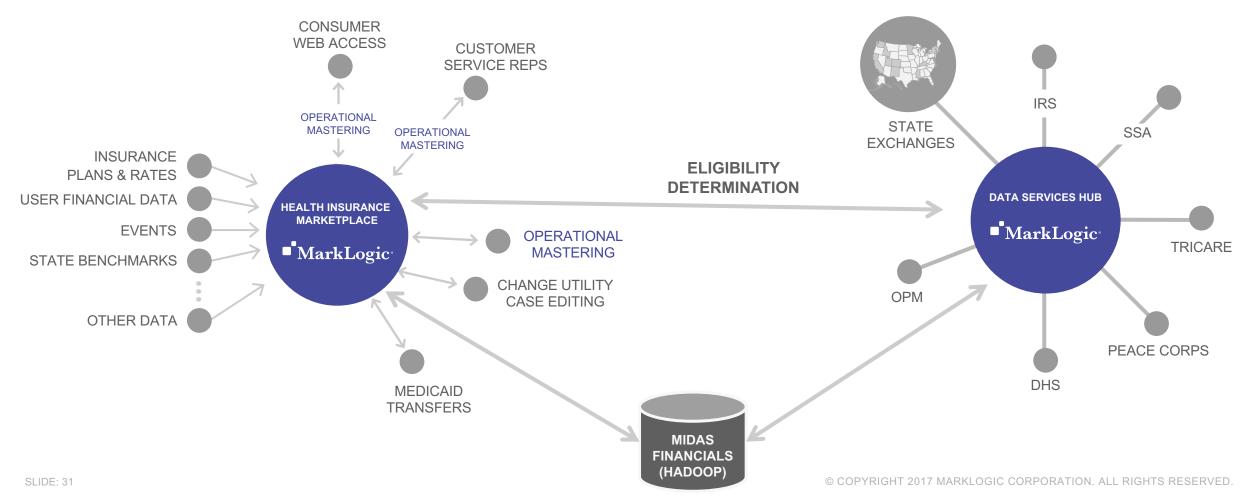
Real-World Examples of Agile Mastering with MarkLogic



HEALTHCARE.GOV

Operational Mastering

Operational Mastering at Point of Engagement Reduces Duplicates and Need for Traditional MDM





CRIMINAL HISTORY SYSTEM

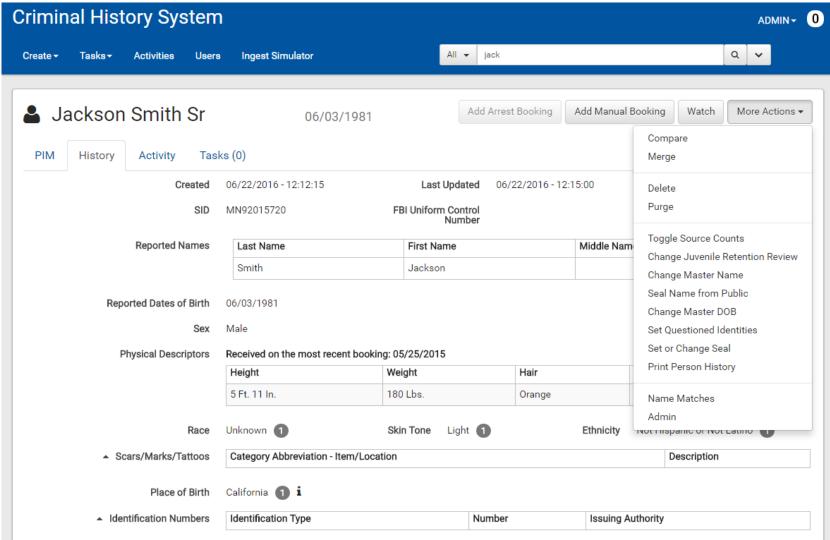
Operational Data Hub with Agile Mastering

Operational Mastering with fuzzy match to find duplicates and correct or merge



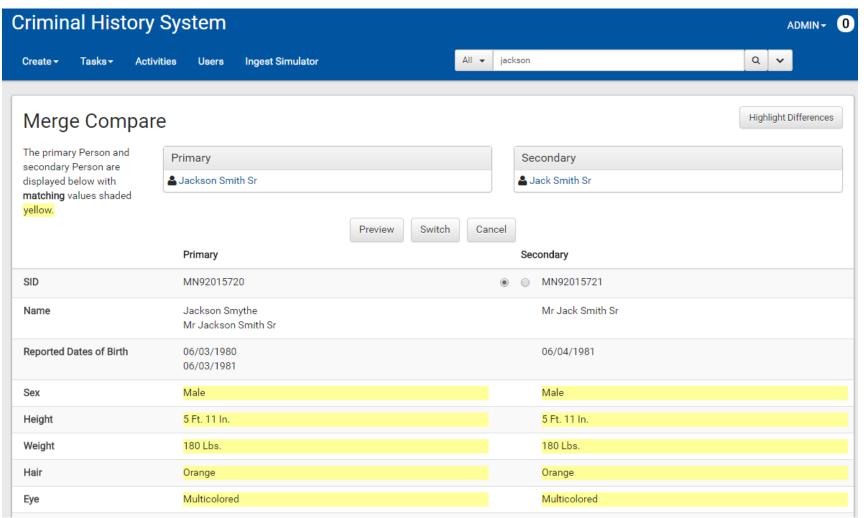


Booking: Ingested Person View





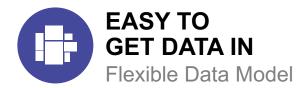
Booking: Merge Compare





The MarkLogic Alternative

An Operational and Transactional Enterprise NoSQL Database Makes Agile Mastering Possible



- Data ingested as is (no ETL)
- Structured and unstructured data
- Data and metadata together
- Adapts to changing data and changing data structures



- Index once and query endlessly
- Real-time and lightning fast
- Query across JSON, XML, text, geospatial, and semantic triples in one database



- Flexible cloud deployment
- Enterprise-grade data security
- Reliable data and transactions (100% ACID compliant)
- Out-of-the-box failover, replication, and backup



Why MarkLogic for Agile Mastering?

VALIDATED

Schema-Agnostic

Contextual

All the Data

Metadata Unlimited

Simplified Architecture

Logical Data Model

Semantics

Secure

Full Auditing



















■ MarkLogic •

Questions?