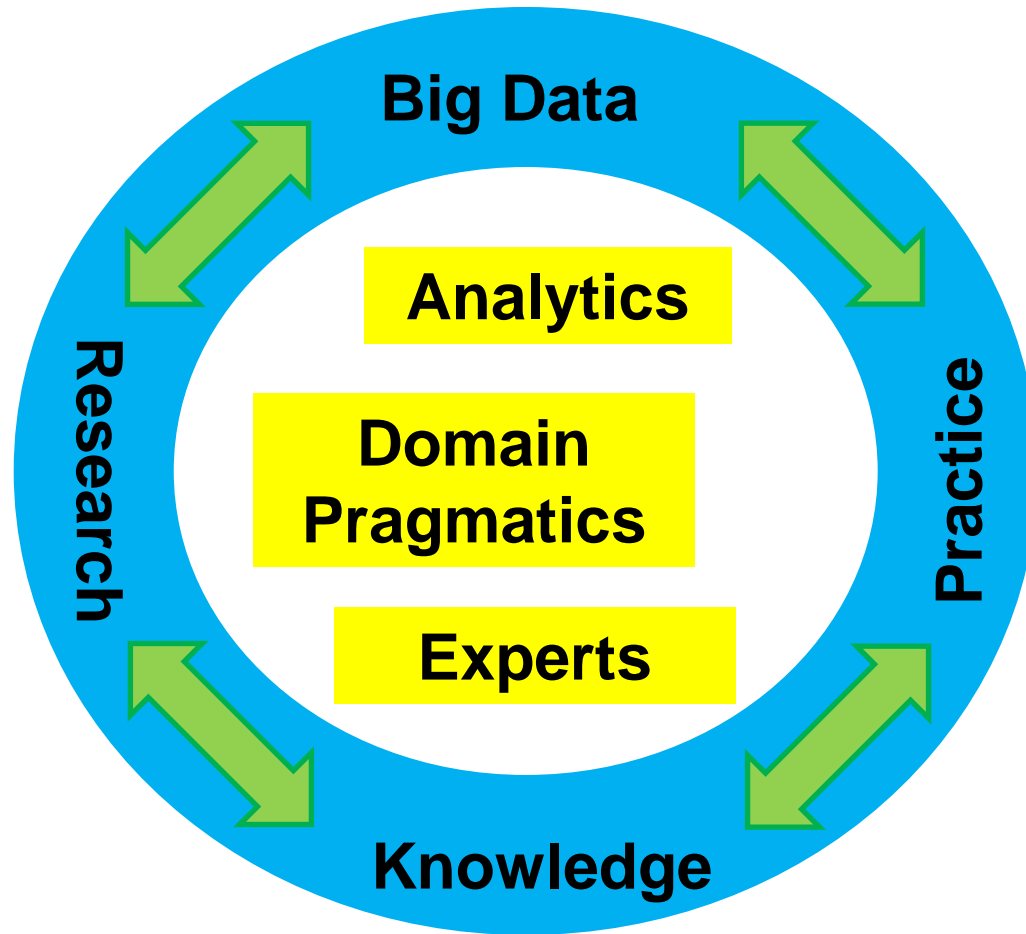


Data Normalization

Hongfang Liu

Data-driven Healthcare

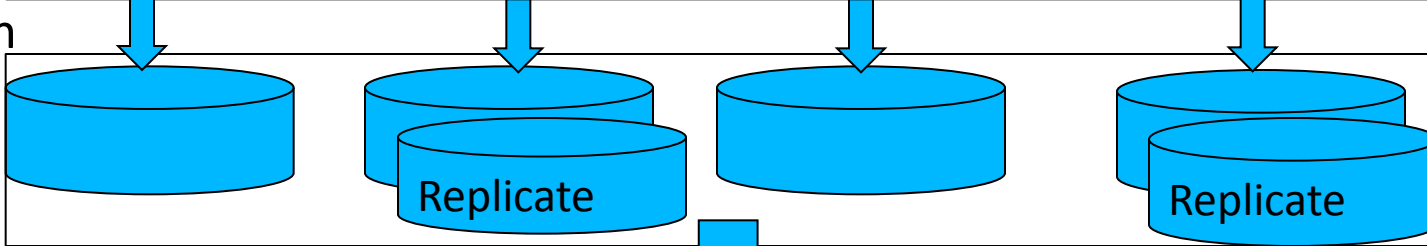


A framework for secondary use of EMRs

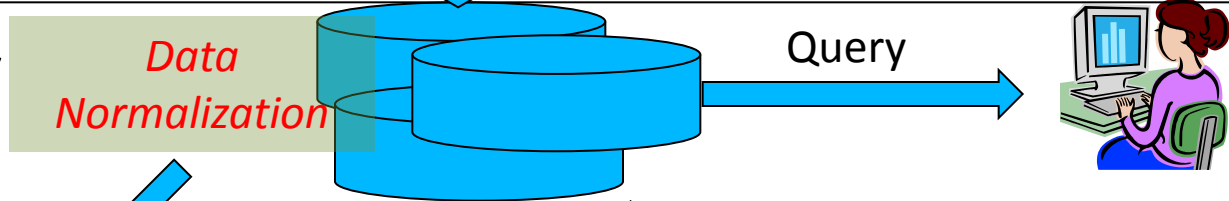
Production Systems



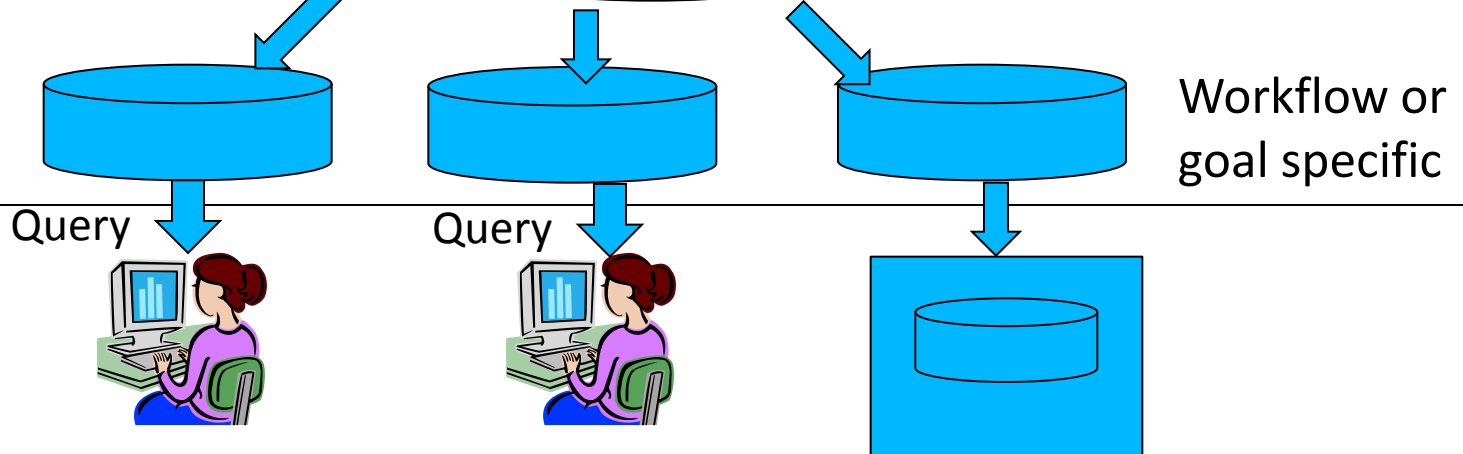
Production Databases



Enterprise Repository/
Data Warehouse



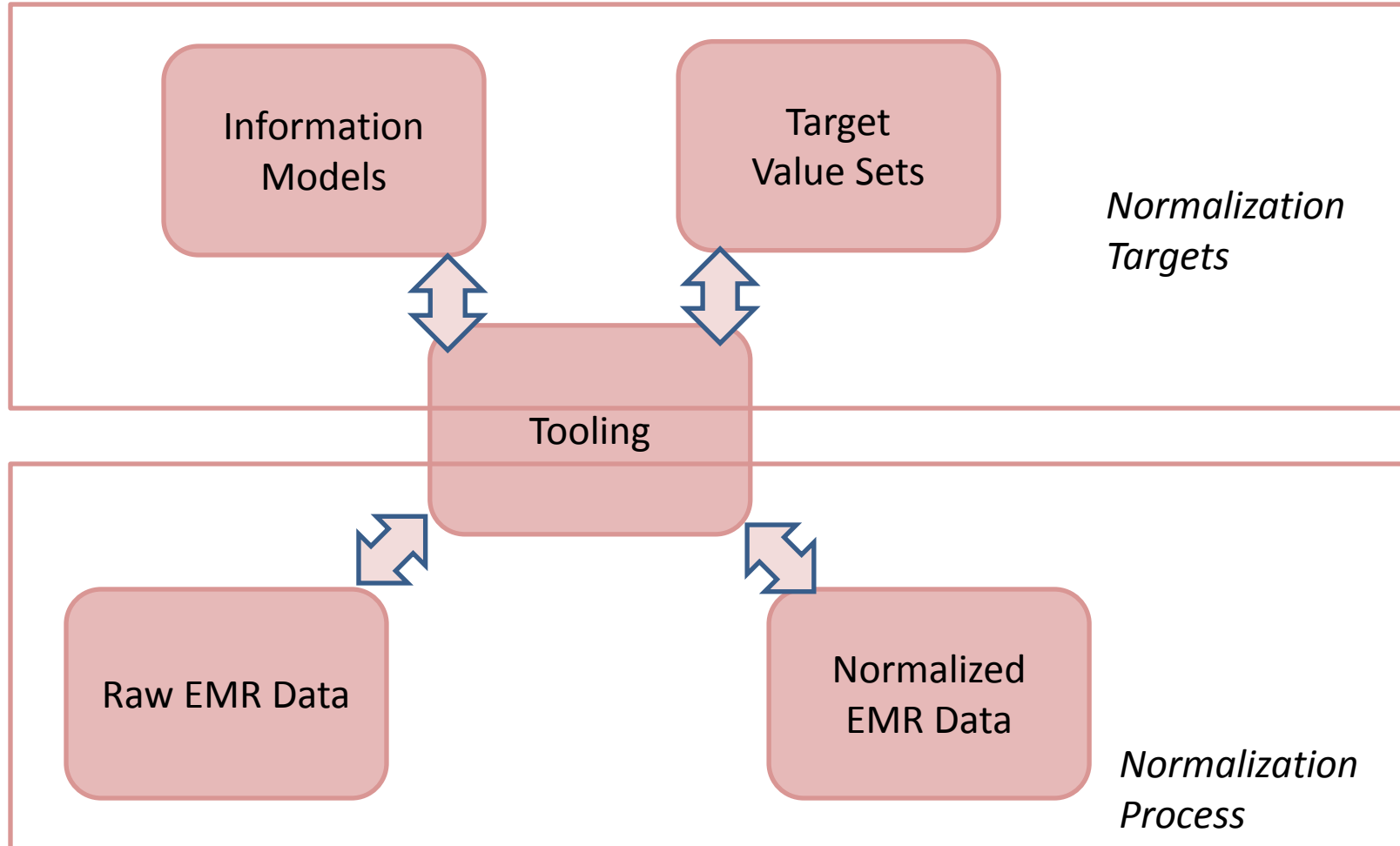
Workgroup
Datamarts



SHARPN DN Goals

- To conduct the science for realizing semantic interoperability and integration of diverse data sources
- To develop interoperable tools and resources enabling the generation of normalized EMR data for secondary uses

Data Normalization



Normalization Targets

- Clinical Element Models
 - Based on Intermountain Healthcare/GE Healthcare's detailed clinical models
- Terminology/value sets associated with the models
 - Using standards where possible

Normalization Process

- Configuration of Model (Syntactic) and Terminology (Semantic) Mapping
- UIMA Pipeline to transform raw EMR data to normalized EMR data based on mappings

Four Subprojects

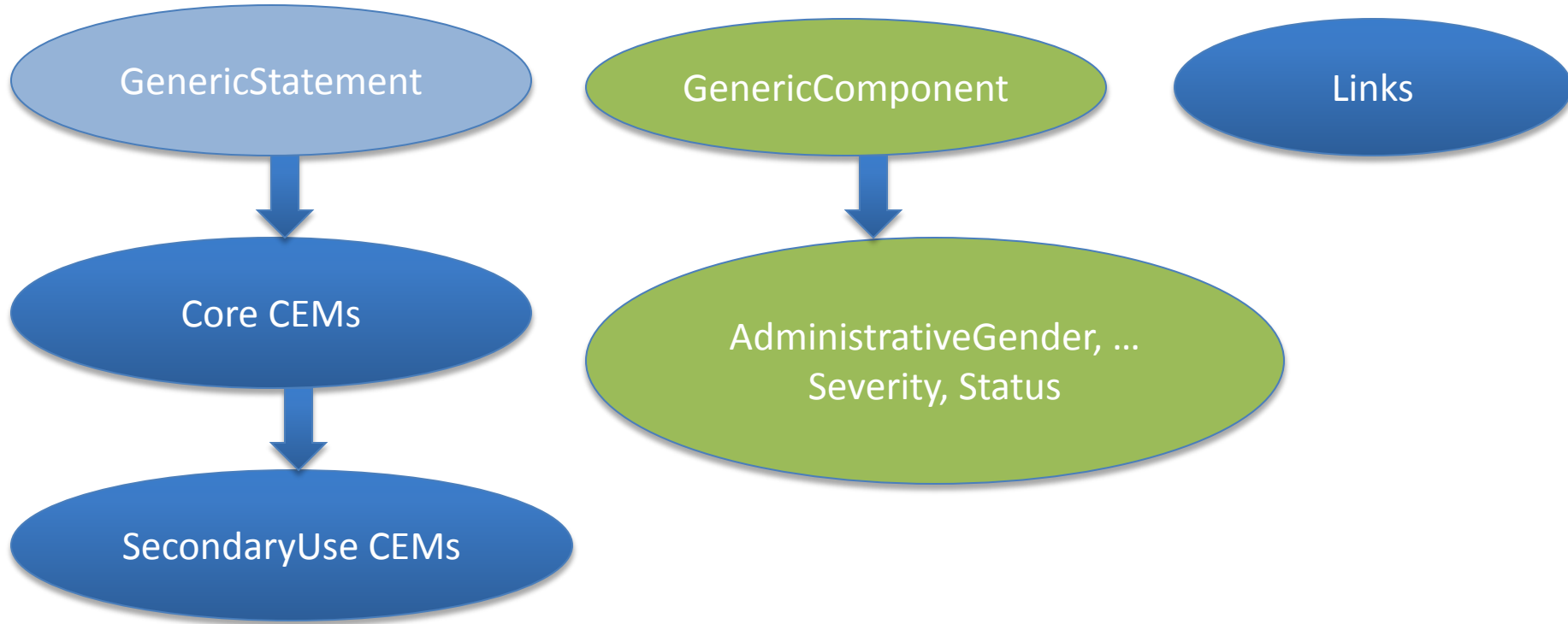
- Clinical Information Modeling
- Value Sets Management (CTS2)
- End-to-End Pipeline
- Normalized Data Representation and Store

Four Subprojects

- Clinical Information Modeling
- Value Sets Management (CTS2)
- End-to-End Pipeline
- Normalized Data Representation and Store

Secondary Use Clinical Element Models

<http://www.clinicalelement.com>



Embracing the fact that data may not be able to be normalized and enabling bottom-up and top-down

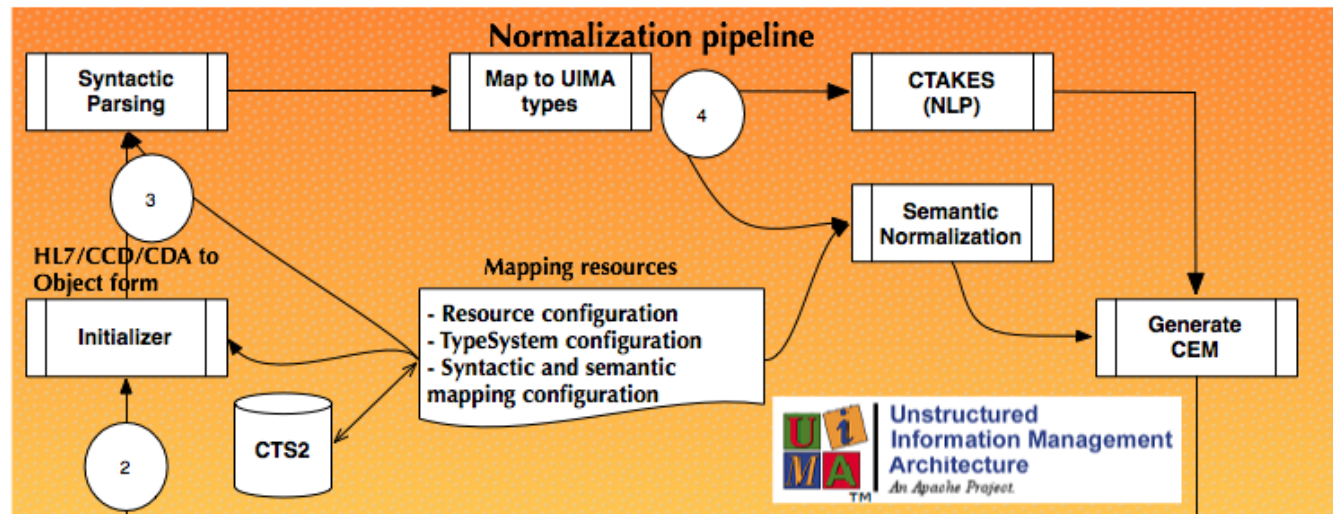
Status of Secondary Use CEMs

- Model specification is final
- CEM Browser is in production
- Manuscript is in preparation

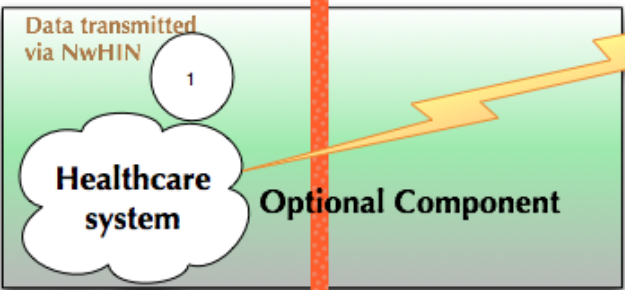
Future:

Secondary Use CEMs and CEM Browser will be maintained through Clinical Information Modeling Initiative (CIMI)

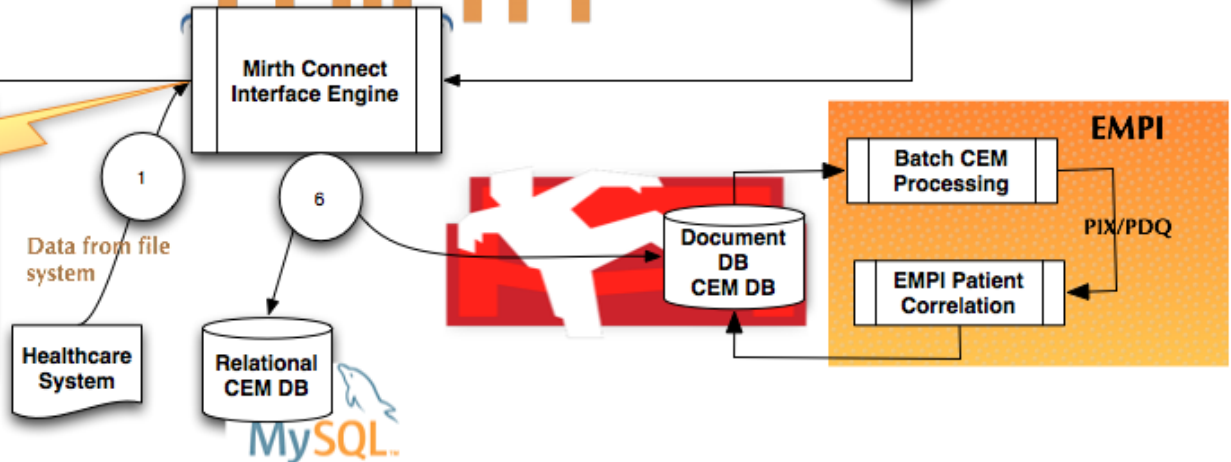
SHARPN Data Normalization Architecture



Firewall



mirth™





Collection Reader

Descriptor:

Browse..

Input Directory:

Browse..

Language:

>>

 SHARPN Data Normalization PipelineEMANTIC_MAPPING: DOCTYPE: CROSSWALKTABLE: Output Directory: Element Property: DEID:

CAS Consumers

Add



Initialized

SecondaryUseNotedDrug – Output (1/2)

```
<CD>  
<code>  
  <value>24941</value>  
</code>  
<codeSystem>  
  <value>2.16.840.1.113883.6.88</value>  
</codeSystem>  
<codeSystemVersion>1.0</codeSystemVersion>  
<originalText>No original text</originalText>  
<translation>  
  <code>FDB~000230806</code>  
  <codeSystem>RXNORM</codeSystem>  
  <codeSystemVersion>Feb 2nd 2012 RXNORM</codeSystemVersion>  
  <originalText>ferrous fumarate-vit C</originalText>  
</translation>  
</CD>
```

destination code

SecondaryUseNotedDrug – Output (2/2)

```
<formulation usage="qualifier">  
  <CD>  
    <code>  
      <value>10311</value>  
    </code>  
    <codeSystem>  
      <value>2.16.840.1.113883.6.88</value>  
    </codeSystem>  
    <codeSystemVersion>Feb 2nd 2012 RXNORM</codeSystemVersion>  
    <originalText>TAB</originalText>  
  </CD>  
</formulation>
```

NLP in data normalization

- A large amount of clinical information is in clinical narratives, NLP is a critical component in data normalization
- cTAKES has been wrapped into the data normalization pipeline to normalize data in clinical narratives

End-to-end DN framework Status

Page Discussion edit History Move Watch

Go Search

About
Main Page
Announcements
SHARPN Program
Who we are
Press
Annual Gathering
Contact Us

Education & Tools
Program Areas
Terminologies and CTS2
Tools & Downloads

Program Administration
Contacts
Meeting Schedule
Meeting Notes
PAC
Reports
Library
Glossary

Projects
1. Data Normalization (DN)
2. Natural Language Processing (NLP)
3. High Throughput Phenotyping (HTP)

Data Normalization Framework 2.0

[Main Page](#) > [Program Areas](#) > [Data Normalization](#) > [Data Normalization Framework 2.0](#)

Overview [\[edit\]](#)

Normalizing data within a healthcare environment means taking all or parts of electronic healthcare documents and transforming them into data that has structure and uses standard underpinnings for terms and measurements and so on. This documentation will refer to this as the Data Normalization framework. You may see it also referred to as a pipeline. It's not like a pipeline you normally think of which is used to transport something. It's more like a manufacturing plant that consumes materials and makes something else.

If you have not done so please study our [thoughts and methodology on data normalization](#). Data normalization involves, on the surface, several parts of an application:

- The incoming data
- A transformation of the data
- The terminology services
- A place to store the results

The release of the end-to-end DN framework is planned on October 1, 2013

This can be accomplished a number of ways. Specialized software is added to process incoming data and pass it to the appropriate processing pipeline. [Clinical Element Models](#) (CEMs) are used as the models to store the pipeline results. From there CEMs can be turned into any number of forms. In 2.0 we are placing them into a CouchDB. Previous releases used a MySQL database.

Out of the box, the Data Normalization pipeline has a limited number of healthcare documents that can be presented as data; they are:

1. [HL7](#) messages
2. [CCD](#) (Continuity of Care Document)
3. [CDA](#) (HL7 Clinical Document Architecture)
4. [XML](#) table (SHARP defined)
5. (being evaluated) [cCDA](#) (HL7 Consolidated Clinical Document Architecture)

[CCDs](#) and [CDAs](#) are XML documents already but [HL7](#) messages are not. Since the pipeline transforms XML, the [HL7](#) messages are treated first and changed into XML. This is one of the functions of Mirth Connect. In essence, the Data Normalization pipeline's syntactic processing step

CEM Representation and Store

- Natively XML document
- Relational database or document database?
- Release 2.0 will be in couchDB
- Implemented: RDF and JSON
- Planned: rendering CCDA documents from CEM instances through OHT tooling

In summary

- Secondary use data model
- CTS2
- Normalization framework
- Data representation and store