Open Source Data Mapping based on RESO Standards
WHY AM I STILL HERE?
• How realtor.com maps your data today
• Modernizing our Listing Processing Platform
  ○ Logic & Workflow for our new system
• How we leveraged RESO’s Data Dictionary to reduce implementation times
• Discuss & Demo an open source ETL Tool that anyone can use
Alex Makhanko
MLS Engineering Mgr
realtor.com
Alex.Makhanko@realtor.com
Update Frequencies

- 51% 1 Minute Updates
- 47% 15 Minute Updates
- Other

#RESO17
- Property Type based mapping

<table>
<thead>
<tr>
<th>MLSs</th>
<th>Data Maps</th>
<th>Total Maps</th>
</tr>
</thead>
<tbody>
<tr>
<td>750</td>
<td>4-6</td>
<td>3,750</td>
</tr>
</tbody>
</table>
Current State

- Property Type based mapping
- Maintenance is rough...

<table>
<thead>
<tr>
<th>MLSs</th>
<th>Data Maps</th>
<th>Total Maps</th>
</tr>
</thead>
<tbody>
<tr>
<td>750</td>
<td>4-6</td>
<td>3,750</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5,000</td>
</tr>
</tbody>
</table>
Current State

- Property Type based mapping
- Maintenance is rough
- Setup takes too long
- Doesn’t leverage RESO Data Dictionary well

**MLSs** | **Data Maps** | **Total Maps** | **Unique Fields** | **Bus Rules**
---|---|---|---|---
750 | 4-6 | 3,50 | 33,000 | 24,000

5,000
“The best time to plant a tree was 20 years ago. The second best time is now.”
- Chinese Proverb

Objectives:
- Turn Days into Hours(!)
- Consolidate rules / simplify
- Cloud-based (AWS)
- Deeper understanding of data
  ○ Empower Search
WHAT WE’VE ACHIEVED SO FAR

- Reduced New implementation Time-to-Market by 70%!
- Leveraged RESO DD to make that happen
- Consolidated 24,000+ mapping rules down to ~ 200 Global rules
- All on an Open Source Cloud-based platform that anyone can duplicate
What is Pentaho

Pentaho Open BI Suite

- Presentation Layer: Portal, Office, Web Services, E-mail
- Reporting: Production, Operational, Ad-hoc
- Analysis: Data Mining, OLAP, Drill & Explore
- Dashboards: Metrics, KPIs, Alerts
- Process Management: Integration, Definition, Execution
- Business Intelligence Platform: Administration, Business Logic
- Data & Application Integration: ETL, Metadata, EII
- Security
- Repository
- 3rd Party Applications: ERP/CRM, Legacy Data, OLAP, Other Applications, Local Data
What is Pentaho

● Open Source ETL Tool
● Connectors to support many data sources (extensible)
● Incredible set of B.I. tools
● Everyone’s an Analyst - Data Sparks Ideas
● Support for Cloud and big data technologies
  ○ AWS, ElasticSearch, Docker, Hadoop
Basic Workflow

1. MLS to RESO Mapping
2. Global Layer (based on RESO DD)
3. Vendor specific layer
   - Corelogic
   - FBS
4. MLS specific layer
   - MLS A
   - MLS B
   - MLS C
5. Quality Layer

- Translate MLS data field names to Data Dictionary
- RESO compliant MLSs greatly reduce reconvert effort
Step One: Match RESO DD!

- Translate MLS data field names to Data Dictionary
- RESO compliant MLSs greatly reduce reconvert effort
Step Two: Global Rules

- Data typing, formatting, ending zeros, cleansing...
- RESO compliant MLSs greatly reduce reconvert effort
- Reduced rules from 36,000 to <200!!
Vendor Specific Rules

Examples:

- FBS - extended remarks in a separate class
- Corelogic/Matrix - room and unit information in a separate class
- Easy to maintain when vendors make changes
MLS Specific Transformation

1. MLS to RESO Mapping
2. Global Layer (based on RESO DD)
3. Vendor specific layer
   - Corelogic
   - FBS
4. MLS specific layer
   - MLS A
   - MLS B
   - MLS C
5. Quality Layer

Search/ F.E.
Data Quality Layer

1. MLS to RESO Mapping
2. Global Layer (based on RESO DD)
3. Vendor specific layer
   - Corelogic
   - FBS
4. MLS specific layer
   - MLS A
   - MLS B
   - MLS C
5. Quality Layer

Search/F.E.

#RESO17
• Series of algorithms that seek anomalies and trends
• Capture questionable records
• Prevent bad data from infecting downstream systems
DEMO: From Days to Hours

- Demonstrate transformations and data flow for a RESO platinum certified MLS using Pentaho Open Source ETL
- Input:
  - MLS data sample data from CRMLS
  - RETS client and Datasource have been pre-setup
- Process
  - Local instance of Pentaho ETL tool called Kettle
  - Application of RESO standard mapping
  - Application of all rules
- Output:
  - Processed data
  - Simulated feature breakdown
- https://meet.google.com/rzz-vsbj-zty
Don't forget to take the session survey in the conference app!