



Search In-Depth

PDS Technical Session
Sean Hardman

February 14, 2018



Topics



- Overview
- Mapping Flow
- Harvest Configuration
- Search Core Configuration
- Search Service Configuration
- Wrap Up

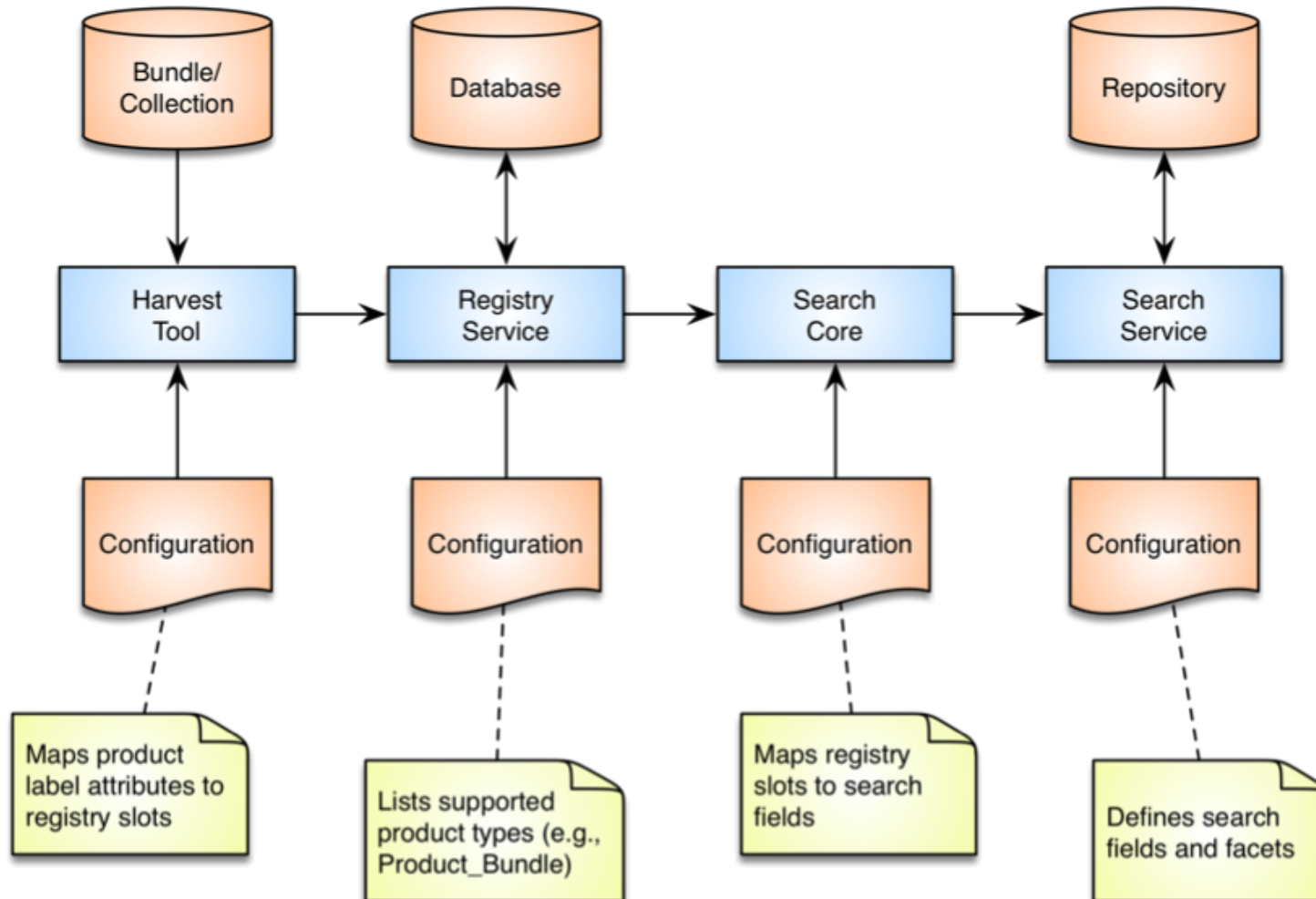


Overview

- Search is driven by a series of attribute mappings as the metadata makes its way through the services
- Although we are “mapping”, we have attempted to keep the attribute names as consistent as possible



Metadata Mapping/Flow





Harvest Tool Configuration

- This is where product label attributes get mapped into Registry slots
- The global policy for the tool includes default mappings for operational, context, bundle and collection products
- Default policy is provided for products like browse, document, observational and XML schema
 - The assumption is that these mappings will likely require tailoring for Node deployments
 - Specifically in the Mission and Discipline Areas



Harvest Tool Configuration Example



```
...
<productMetadata objectType="Product_Observational">
  <!-- Identification_Area -->
  ...
  <!-- Observation_Area -->
  <xPath slotName="observation_start_date_time">
    //Observation_Area/Time_Coordinates/start_date_time
  </XPath>
  <xPath slotName="observation_stop_date_time">
    //Observation_Area/Time_Coordinates/stop_date_time
  </XPath>
  ...
  <!-- Mission_Area -->
  <xPath slotName="latitude">
    //Observation_Area/Mission_Area/ladee:latitude
  </XPath>
  <xPath slotName="longitude">
    //Observation_Area/Mission_Area/ladee:longitude
  </XPath>
  ...
</productMetadata>
...
```



Reference Handling

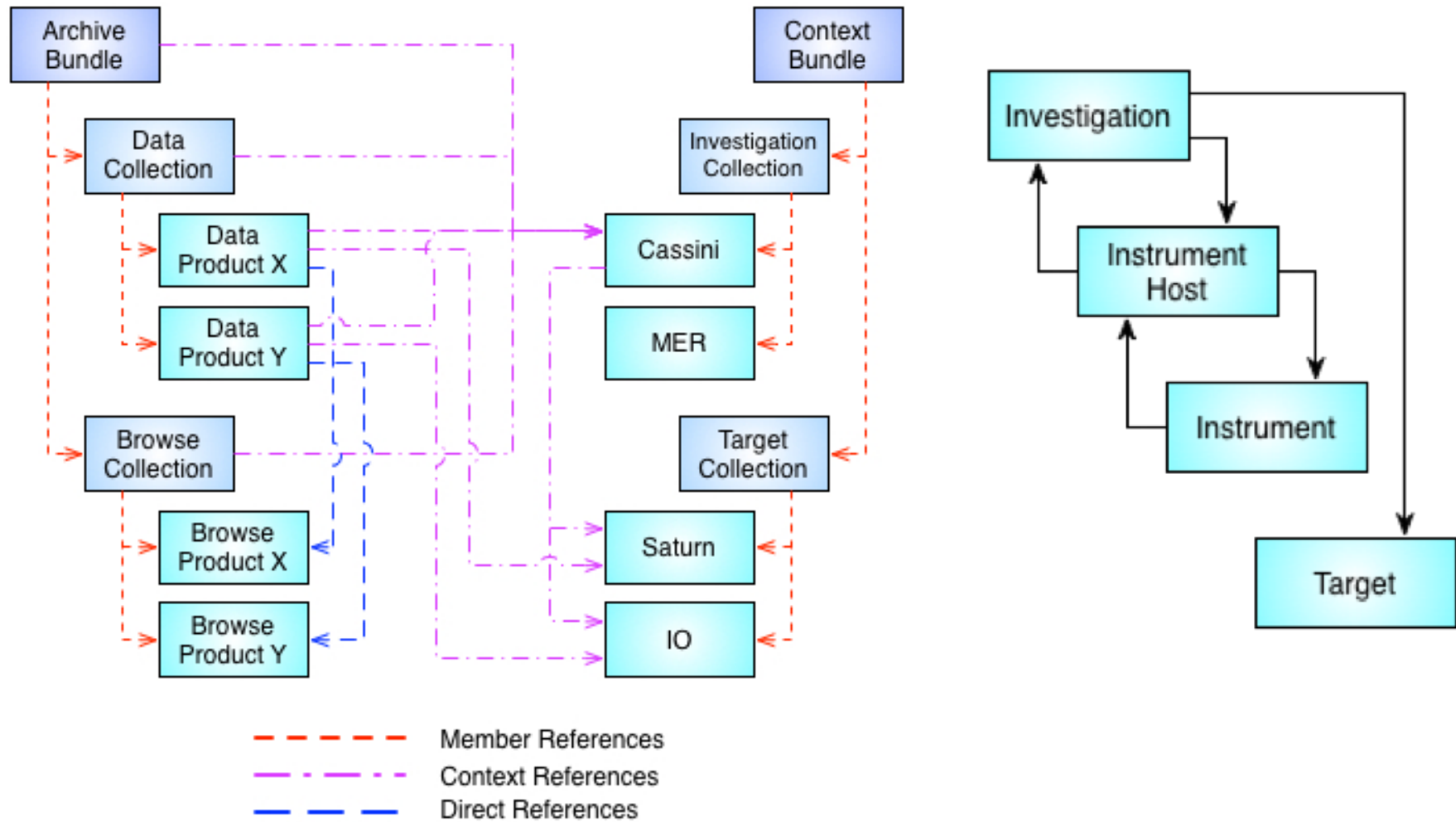
- Internal references are handled internally by the Harvest Tool
- A reference block is mapped to a specific slot based on its reference type

```
<Internal_Reference>  
  <lid_reference>...:target:planet.jupiter</lid_reference>  
  <reference_type>data_to_target</reference_type>  
</Internal_Reference>
```

- The above reference is mapped into a slot named “target_ref” with a value of the LID/LIDVID



References





Search Core Configuration

- This is where Registry Slots get mapped into Search fields
- Like Harvest, the tool comes with a default policy
 - The exception is that we only index a subset of the products
- This configuration drives the generation of the Solr search index
 - This is done in three steps: extract, index and post



Search Core Configuration Example



```
<product>
  <specification>
    <title>PDS4-Observational</title>
    ...
  </specification>
  <indexFields>
    <!-- Observation_Area Fields -->
    <field name="observation_start_date_time" type="date">
      <registryPath>observation_start_date_time</registryPath>
    </field>
    <field name="observation_stop_date_time" type="date">
      <registryPath>observation_stop_date_time</registryPath>
    </field>
    ...
    <!-- LADEE Fields -->
    <field name="ladee_latitude" type="float">
      <registryPath>latitude</registryPath>
    </field>
    <field name="ladee_longitude" type="float">
      <registryPath>longitude</registryPath>
    </field>
    ...
  </indexFields>
</product>
```



Search Service Configuration

- This is where the search fields and facets are defined
- Search fields include direct mappings from the registry slots as well as fields used by the search protocol (e.g., identifier, instrument, etc.)
- Up until this point, values were captured and mapped as simple strings
 - Solr allows us to distinguish between data types where appropriate as well as how a string will be indexed and searched



Search Service Configuration



Example

```
<schema name="pds" version="1.4">
  ...
  <field name="observation_start_date_time" type="date"
indexed="true" stored="true" multiValued="true" />
  <field name="observation_stop_date_time" type="date"
indexed="true" stored="true" multiValued="true" />
  ...
  <!-- LADEE UVS Fields -->
  <field name="ladee_latitude" type="float"
indexed="true" stored="true" multiValued="false" />
  <field name="ladee_longitude" type="float"
indexed="true" stored="true" multiValued="false" />
  ...
</schema>
```



Wrap Up

- Once the products are indexed, they are served up by multiple end points from the service
 - End points are configured with augmentations to the query to serve a specific need (i.e., hide superseded data sets)
- Search interfaces offer additional tailoring for specific products

Questions/Comments