

From Cataloging to User Search – the path to Primo

Presented by

Lynn Wolf

ODIN Technical Support

ODIN Work Day 2016

April 19, 2016

Introduction

- Have you ever wondered how your record reaches the Primo result set after it is cataloged? In this session you will get a brief glimpse at some of the processes required to make a record searchable in Primo.
- Topics will include data sources, publishing, harvesting, normalization rules, and indexing.

Key Terms

- Data Source
 - Where the records originate
- Publish
 - How the records prepared and packaged
- Harvest
 - How prepared records are collected from the Data Source
- Normalization
 - How records converted to uniformed Primo format – PNX
- Indexing
 - Final step in processing records so they can be searched

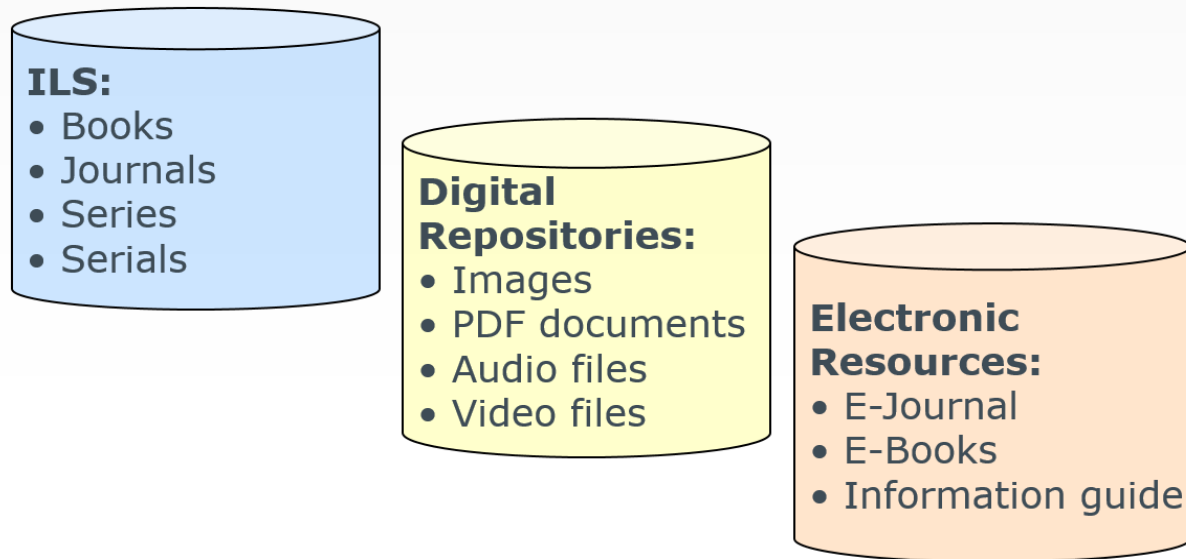
ODIN Primo configuration uses several data sources

- ALEPH
- ALMA
- ContentDM
- DSPACE

Definitions > Data Source

Data Source = Source System

The Primo database may contain data from many different data sources:



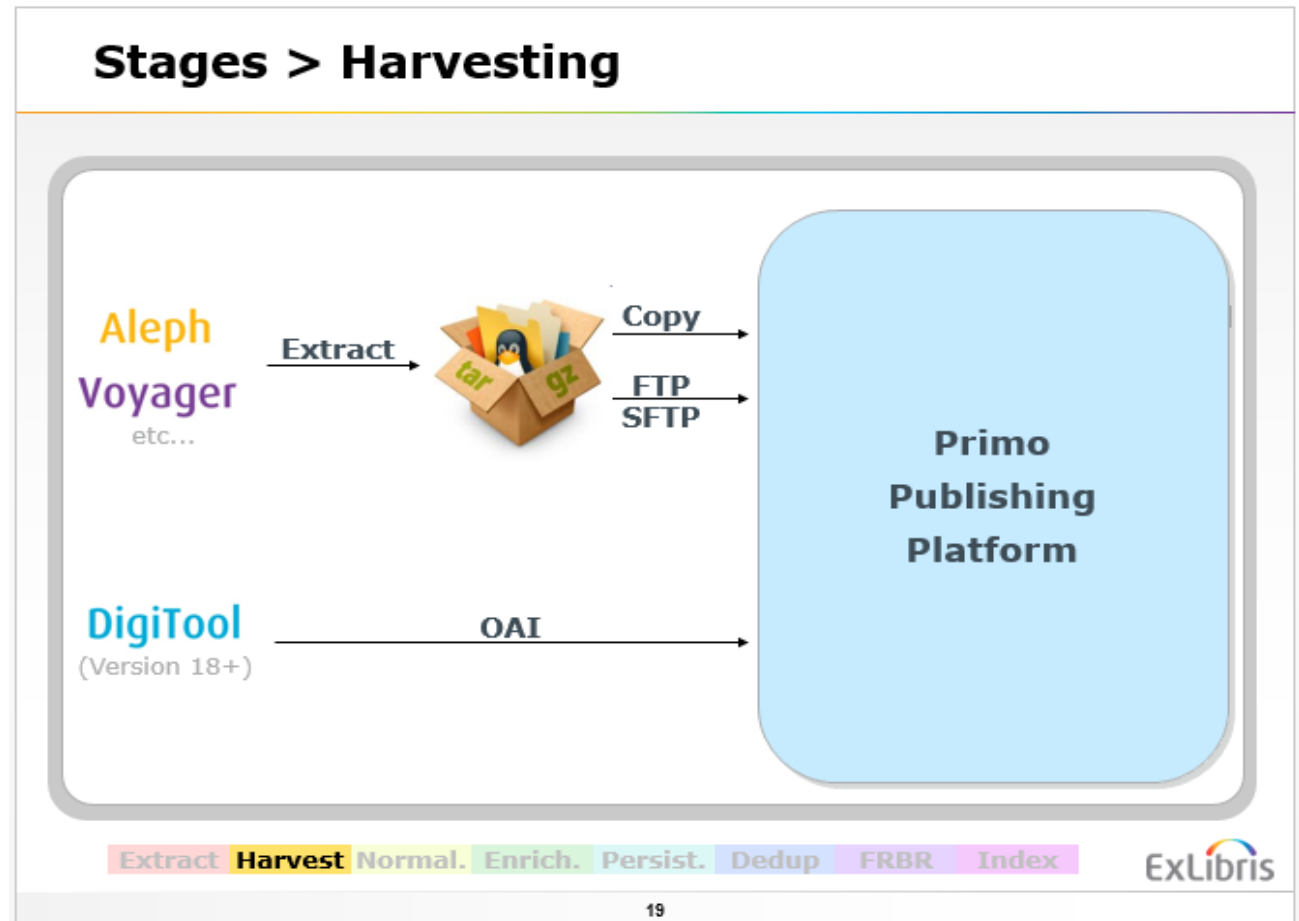
Publishing

- In ODIN ALEPH, the publishing process is run daily in the early morning hours. The process packages the records identified for publishing, applies any defined fix and expand procedures, and then creates a file on the ALEPH server.
- ALMA also has a publishing process that exports new and changed records on a daily basis. ALMA places the created file on a server to be retrieved by Primo.
- Likewise, ContentDM and DSPACE sources have to create files on their host servers for pickup by a Primo process.



Harvesting records

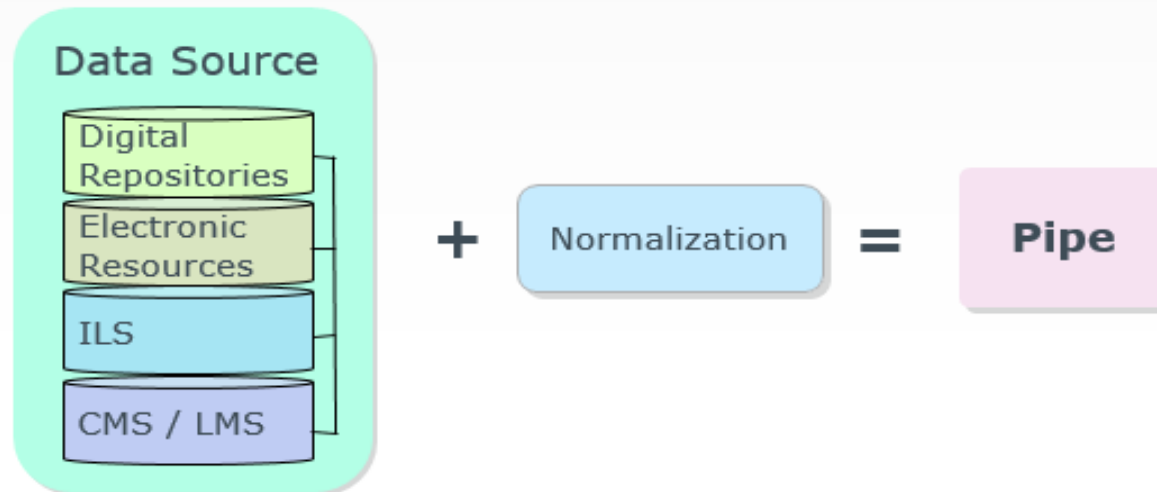
- Primo uses a process called Harvesting to collect files from the different data sources.
- Depending on the source, this process could use FTP, SFTP, file copy or OAI (Open Archives Initiative Protocol for metadata harvesting).
- Each data source has a “pipe” defined specific for that configuration to retrieve the file.
- Each pipe has a schedule configured in Primo to go out and look for new files to process.



Pipe

Definitions > PIPE

A process in Primo by which the source data is converted into PNX format and then loaded into the database.

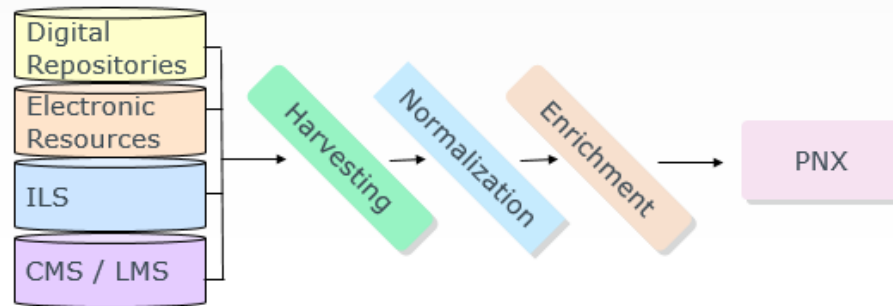


Normalization

Definitions > Normalization

Normalization is one of the main functions of the publishing platform.

Normalization consists of converting data from all sources into a uniformed Primo format – PNX.



More about Normalization (Norm Rules)

- When records come into Primo, they are converted through Norm Rules into the PNX format.
- There are 12 different subfields in the PNX record. Each field and its subfields are used in different aspects of Primo.

Definitions > PNX

Control

Display

Links

Search

Facets

Sort

Dedup

FRBR

Delivery & Scoping

Ranking

Enrichment

Additional Data

```
<?xml version="1.0" encoding="UTF-8" ?>
<record>
  <control>
    <sourceid>primo_aleph</sourceid>
    <recordid>primo_aleph004536586</recordid>
    <sourcesystem>Aleph</sourcesystem>
  </control>
  <display>
    <type>journal</type>
    <title>Wall Street journal (Eastern ed.)</title>
    <edition>Eastern ed..</edition>
    <publisher>New York, N.Y. : Dow Jones</publisher>
    <creationdate>1959-</creationdate>
    <identifier>$$CISSN$$V0099-9660</identifier>
    <subject>New York (N.Y.) -- Newspapers;</subject>
    <language>eng</language>
    <source>primo_aleph</source>
  </display>
  <links>
    <backlink>$$Taleph_backlink$$</backlink>
    <linktoholdings>$$Taleph_holdings</linktoholdings>
  </links>
  <search>
    <title>The Wall Street journal.</title>
    <subject>New York (N.Y.) Newspapers.</subject>
    <general>Dow Jones.</general>
    <issn>0099-9660</issn>
    <issn>0043-0080</issn>
    .
  </search>

```

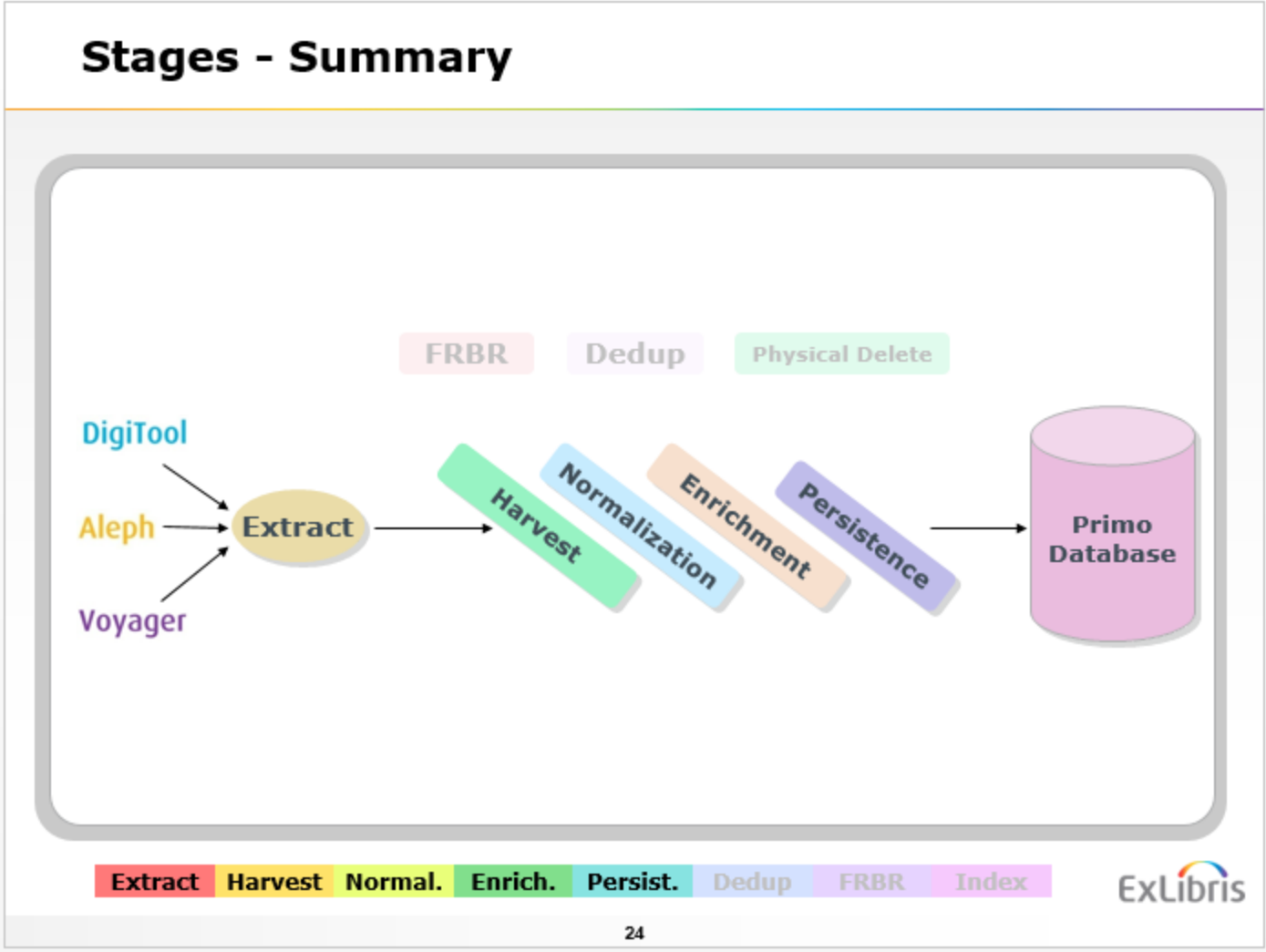
Still more about Norm Rules

- Primo comes with predefined templates (NR mapping sets) that contain the basic configurations recommended for certain data sources.
- Our libraries have started with the basic configurations and made minor adjustments over time.
- Service Packs / Upgrades may also make adjustments.
- When changes are made, the Primo database must be “renormed” and indexed. Depending on which data source is being renormed, the entire process can take anywhere from 7 to 26 hours.
- For the most part, searching on Primo is still available during these renorms as long as the Search Engine is not restarted.

NEP --- Normalization ... and what else?

- Enrichment – where source data can be enriched with different content not originally residing in the source record but coming from other external sources. Primo offers several enrichment programs (such as LCCN call number enrichment). Currently all ODIN pipes use the default template designated for their data source.
- Persistence – where the results of the PNX record is written into the database. It also cleans the database from source records that were marked for deletion.

Let's recap the process so far ...



Indexing

- Indexing is a separate process.
- As records go through the NEP process, they are flagged for indexing.
- After indexing is completed there are some activities that go on “behind the scenes” before you are actually able to search for your newly integrated records.



Questions?





Credit to Alon Botvinik, Ex Libris Primo Support, for many of the slides were taken from his Primo Publishing Platform presentation at the IGeLU Conference & Ex Libris Systems Seminar in Haifa Israel – September 2011.

Thank you for attending!