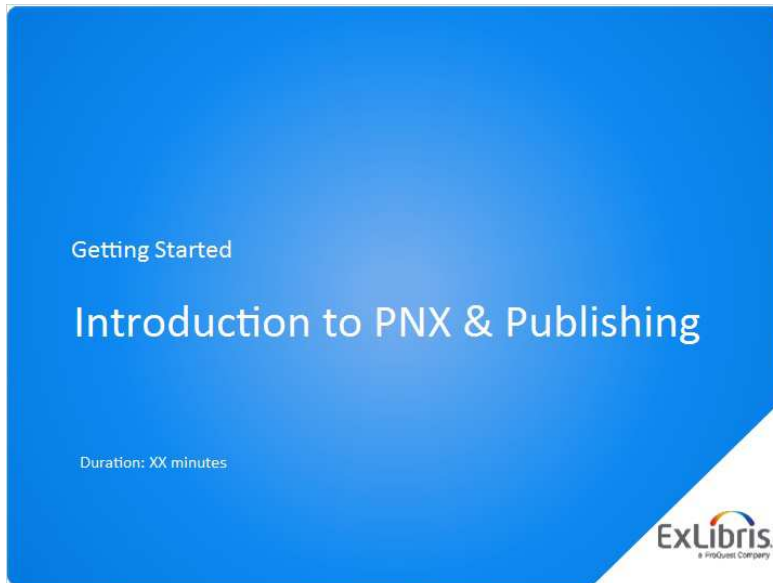# Getting Started: Introduction to PNX and Publishing

1. Introduction to PNX and Publishing

1.1 Introduction to PNX and Publishing

Notes:

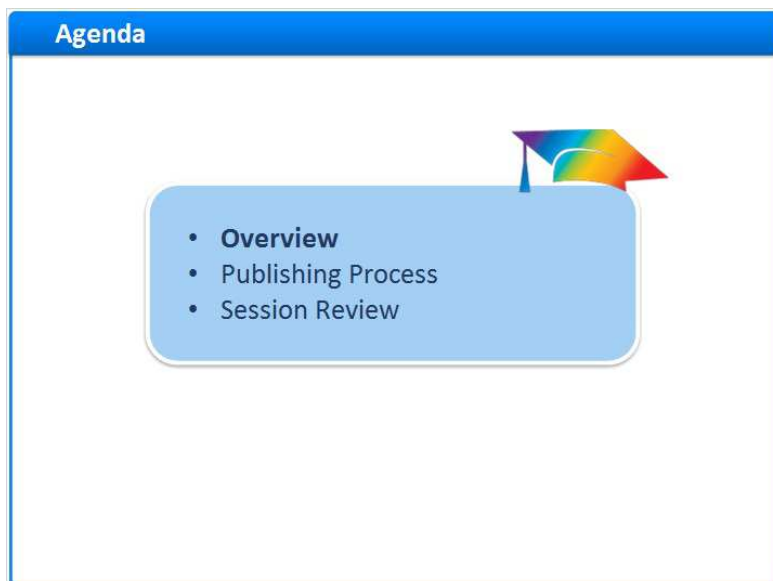Hello, and welcome to this session Introduction to PNX and Publishing

1.2 Agenda

Notes:

We will start with a brief overview of what Primo Normalized XML, or PNX is. Then we will go through the publishing process step by step to see how records are published in Primo.

1.3 Overview

Notes:

Overview

1.4 Introduction

Notes:

Here we have a record in Primo for the book *Diverse methodologies in the study of music teaching and learning*
In the full record, the Details section contains the title, author, subjects, a description, related titles, the publisher, creation date, format, language, identifier and source.
This information is taken from the MARC record coming from the ILS, but when it was loaded into Primo it was transformed from MARC into a different type of record and stored in the Primo database.

1.5 Primo Normalized XML (PNX)

Primo Normalized XML (PNX)

Why do we transform records?

- Start with different data repositories and different formats
- End result – all records in same format
- Allows Primo to search all records at once
- Consistent display
- Scopes, filters, browse

What do records get transformed into?

- Primo Normalized XML (PNX)

Notes:

So why do we need to transform these records?
While it's possible to extract data from library catalogs in various forms of MARC, Primo can harvest data in other formats as well, such as Dublin Core, XML, and the Web Archive format. This allows us to load different types of data into Primo, and have all of the records end up in the same format.
Having all of the records formatted the same way allows Primo to search across all records at once, and display them consistently in the discovery layer. It also allows us to organize and group records in different ways to create search scopes, filters, and enable virtual shelf browsing.
And what format do these records get transformed into? It's a proprietary XML format, called Primo Normalized XML.

1.6 Publishing Process

Notes:

Publishing Process

1.7 Overview of Publishing Process
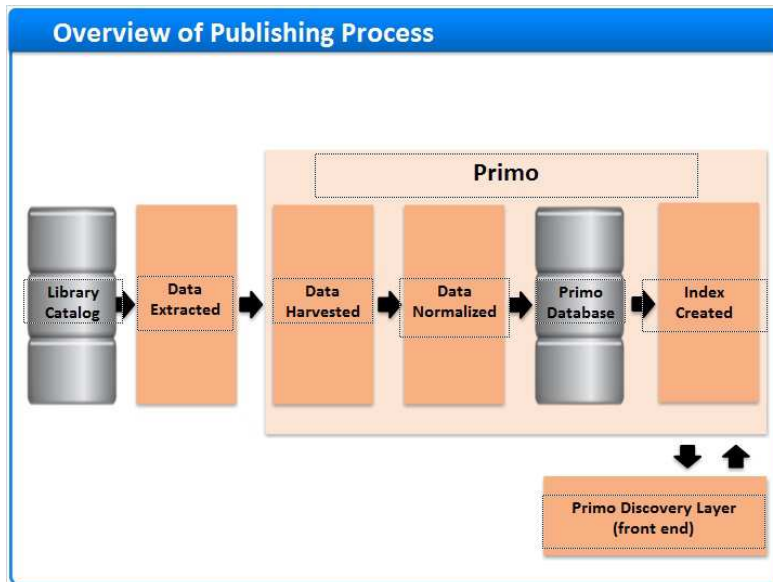
Notes:

So now that we know we're going to be converting your source records into PNX records, let's take a look at how the data is processed from start to finish.

It all starts with your catalog, which already contains your records, most likely in MARC. Data is extracted from your ILS, harvested into Primo, and transformed into the PNX records, or 'normalized'. Once the records are in the Primo database, an index is created. The Primo discovery layer, or 'front end' uses this index to quickly search and retrieve results for your end users.

Now let's take a closer look at each of these stages.

1.8 Publishing - Library Catalog
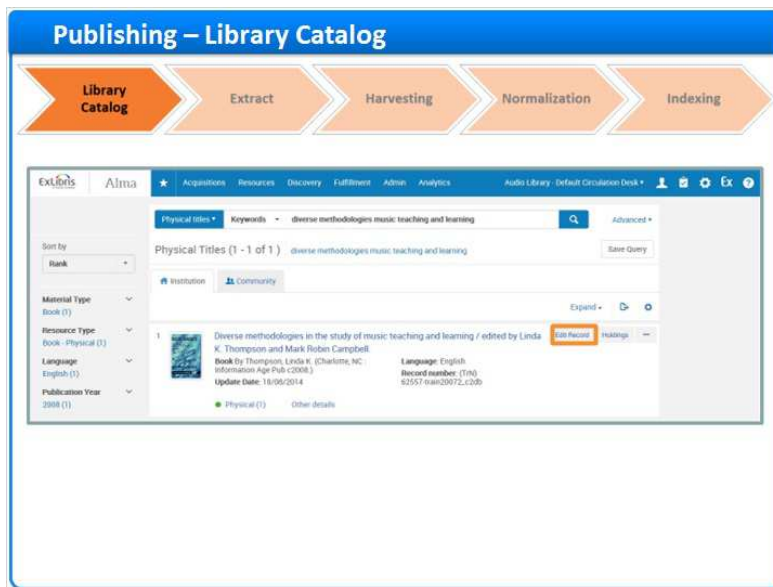
Notes:

Here is our book in the library catalog. There is a brief summary of the resource, and if I select **Edit Record**, I can also see and edit the MARC record.

marc record (Slide Layer)
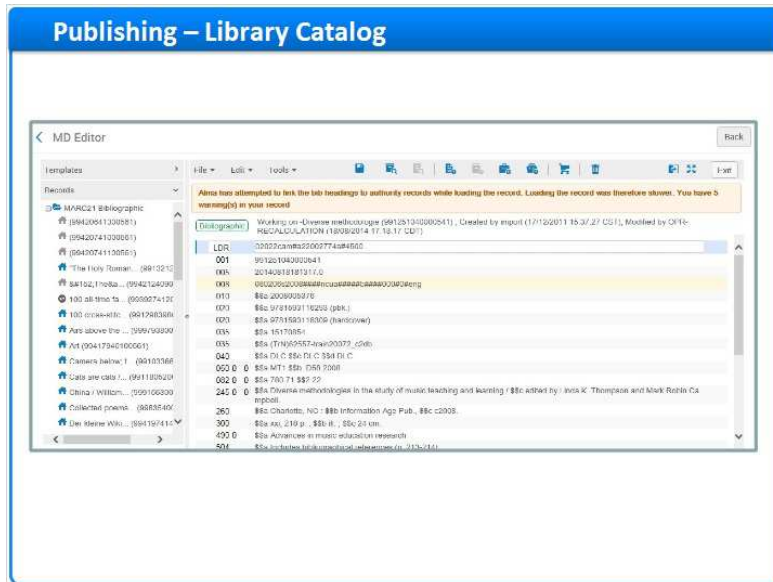
1.9 Publishing - Extract

Notes:

While your records are stored in MARC in the ILS, one of the first steps in your Primo implementation will be to generate an extract of this data.

Primo can process data in many different formats. Our Alma data, for example, will be extracted in MARC21 XML. Our sample record, once extracted would look like this.

We have the 245 field, which contains the title and author information, the 650 field, which contains

---

the subjects, and so on.

One thing to keep in mind about the extract is that each record will include both the MARC record itself, and the availability information, including the library, collection, and location codes associated with this record. In our example here, that information is present in the AVA section.

The first time you generate an extract, it will be for your entire data set. During your implementation, we'll load this initial extract, and, then we'll work with you to set up ongoing extracts to capture changes to your data on a regular basis.

Let's imagine for now that this is a new record, and we're loading it for the first time. The next step would be to harvest this record into Primo.

marc xml (Slide Layer)

1.10 Publishing - Harvesting
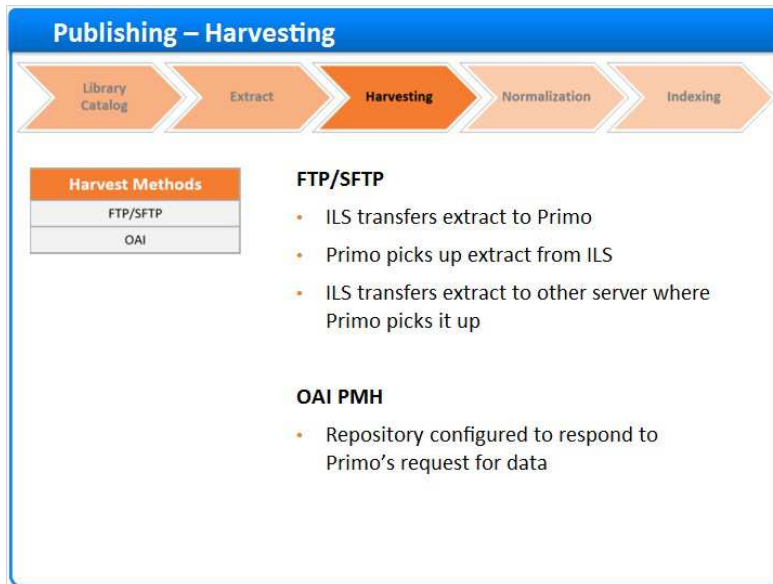
Notes:

In this stage of the process, Primo harvests the extracted records.

Primo can harvest extracts in several ways. While it's possible to manually place the data on the Primo server, during your implementation we'll configure an automated way to transfer these files.

Your ILS may be configured to drop the files onto the Primo server, or Primo can pick up the files from the ILS. It's also possible to configure the ILS to move the files to a different location, and have Primo pick them up there. All of these options use file transfer protocols, or FTP.

OAI protocols are used to harvest data from various digital repositories - instead of generating an extract file, Primo can request the data directly from the OAI server during the harvesting process.

Once we determine the best method to transfer the data, we can schedule a task in Primo to automatically harvest new extracts on a regular basis.
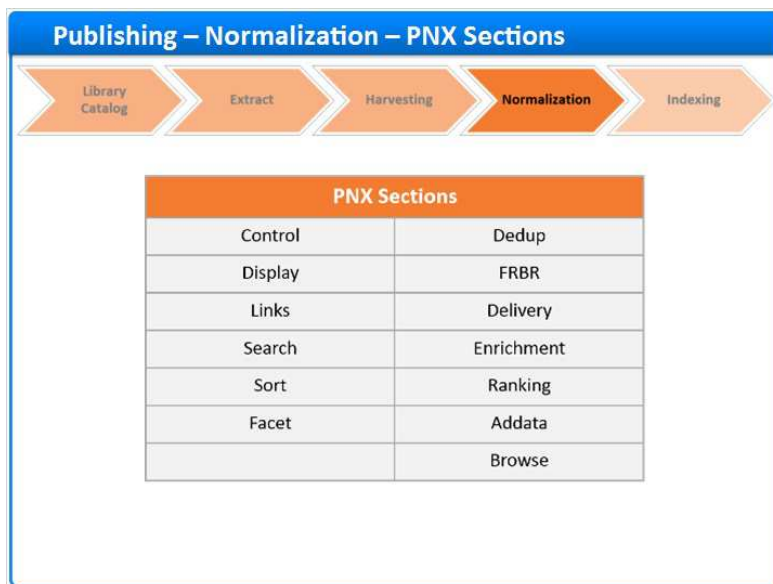
1.11 Publishing - Normalization
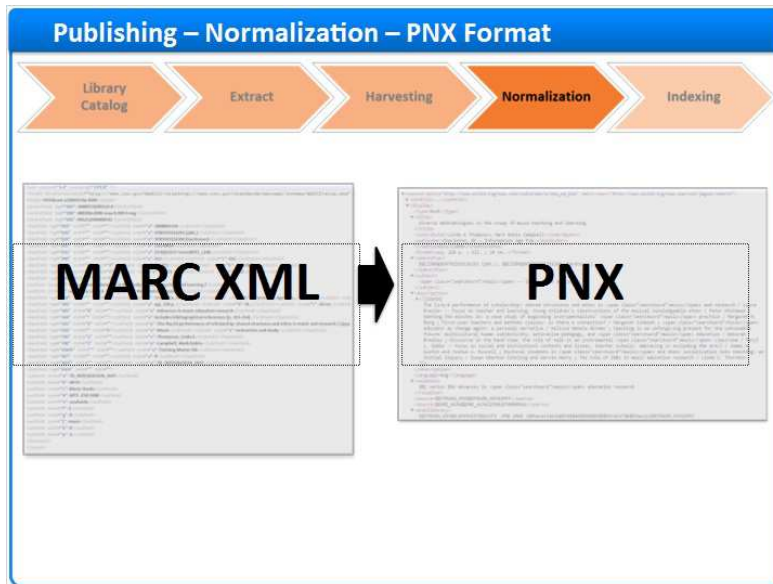
Notes:

After Primo harvests the data, it converts the records into Primo Normalized XML.
The normalization process itself consists of a set of rules that determine how to transform the data from the source record, which in this case is in MARC XML, to the appropriate fields in the PNX record

1.12 Publishing - Normalization

Notes:

Thirteen different sections are created in the PNX during the normalization process. These sections are populated with the data from the original record in various ways. Each section serves a purpose in Primo, whether it's to present the record for display, or optimize it for searching.

1.13 Publishing - Normalization Rule
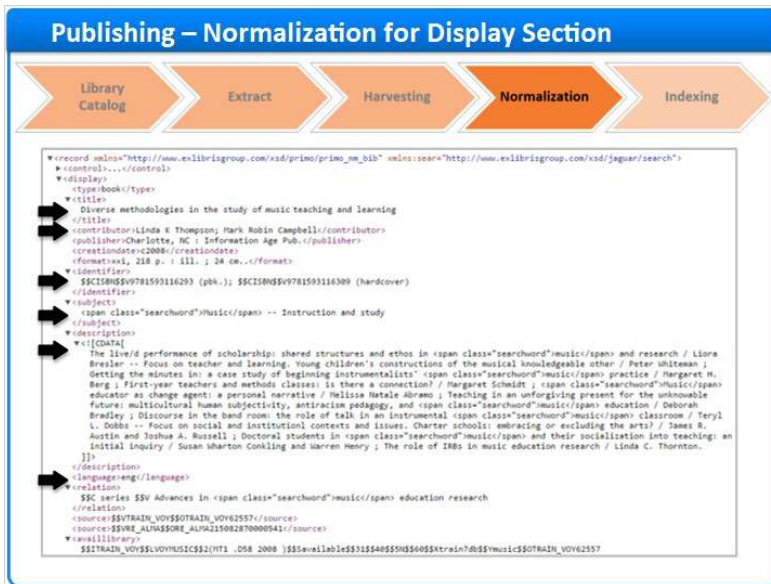

Notes:

To see a basic normalization rule in action, let's go back to our MARC XML record and look at how the title was transformed.

Here we have the title in the MARC, in 245 subfield a. We can see that the title is 'Diverse methodologies in the study of music teaching and learning.' There's also a forward slash at the end of the title.

Next, we have the normalization rule in Primo. We can read the rule like this "Populate the display_title field, by going to the MARC 245 field and taking the data from the a,b,f,g,k,n, and p subfields. Remove these characters (including our forward slash) from the end of the string."

By applying this rule, we're able to generate the title field in the display section in the PNX. You can see here that it took the entire title from 245$a, and since the c subfield wasn't included in the normalization rule it was ignored. The rule also stripped out the forward slash at the end of the string

1.14 Publishing - Normalization for Display Section

Notes:

The Display section in the PNX contains more than just the title - it can hold a number of fields, depending on the data that was present in the original record.

This is the full display section of the PNX for our sample record

We can see all of the fields we want to display in Primo, including the title, contributors, identifiers, subjects, the description, language and more. For each of these PNX fields, there is a corresponding normalization rule that was defined to process the data found in the MARC record.

1.15 Publishing - Normalization for Search Section
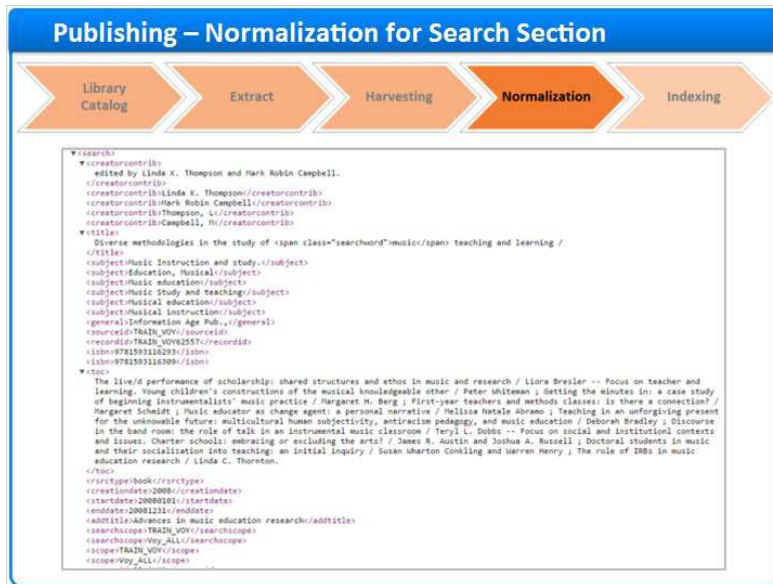
Notes:

Here we have an other section of the PNX.

The search section is a bit different from the display. Some data fields are the same, like the contributor line here, but there are also a few new fields. Notice that we have separate fields for each creator/contributor. Again, the data itself is being taken from the MARC record, notably the 245$c and the 700$a, but Primo is also standardizing and optimizing the content specifically for searching.
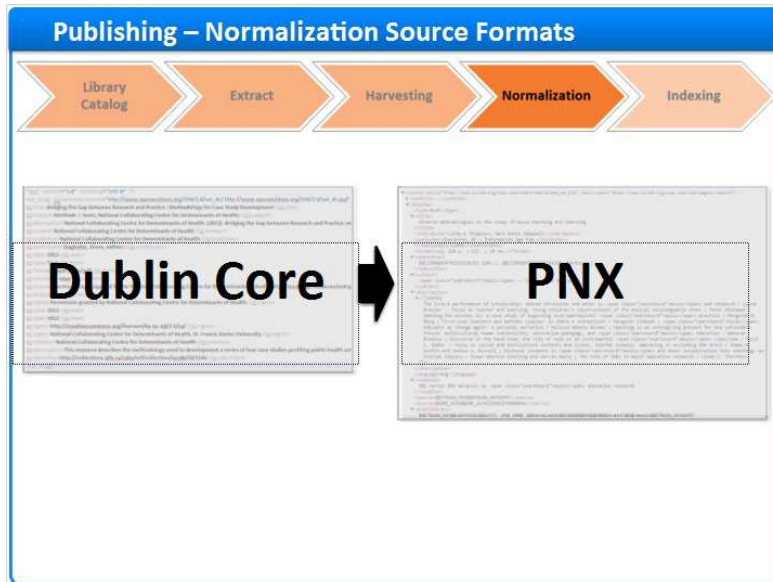
1.16 Publishing - Normalization
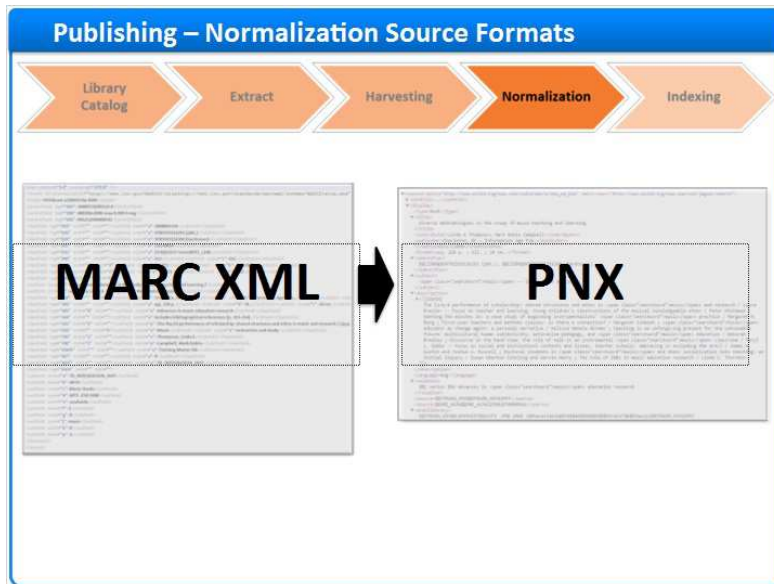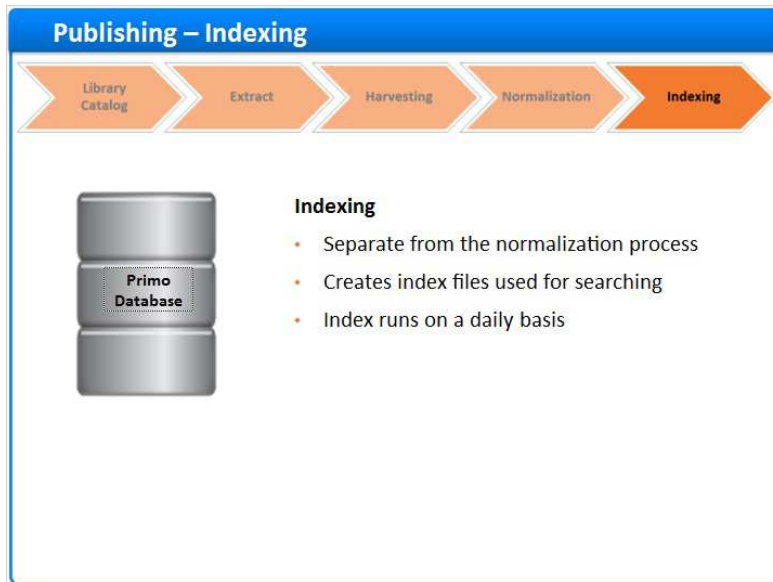

Notes:

While your library catalog extracts are in MARC XML, your digital repository may provide data in Dublin Core format.   Since your sources may provide files in different formats, each data source will have its own set of normalization rules.

Having these separate rules allows us to standardized your records, regardless of the source, into PNX.

More specific information about the normalization rules will be covered in future training sessions.

During your implementation, you'll be looking at your own data after it has been normalized, and working with us to ensure that your local fields are getting written to the PNX correctly.

dublin core (Slide Layer)


1.17 Publishing - Indexing

Notes:

The Primo database itself will have a copy of the record in its original format as well as the PNX record.
Now that we have the normalized PNX records in our database, we are ready for indexing.
The indexing process is separate from the normalization process.
During the indexing process, certain sections in the PNX records are used to generate index files that
are referenced when a user searches Primo. Without the indexing process, the records in the database

are not available to the Primo front end for searching.

Usually indexing is set to run on the Primo server on a daily basis so that it picks up any changes to the data that was harvested and normalized from all of your sources.

1.18 Summary of Publishing Process
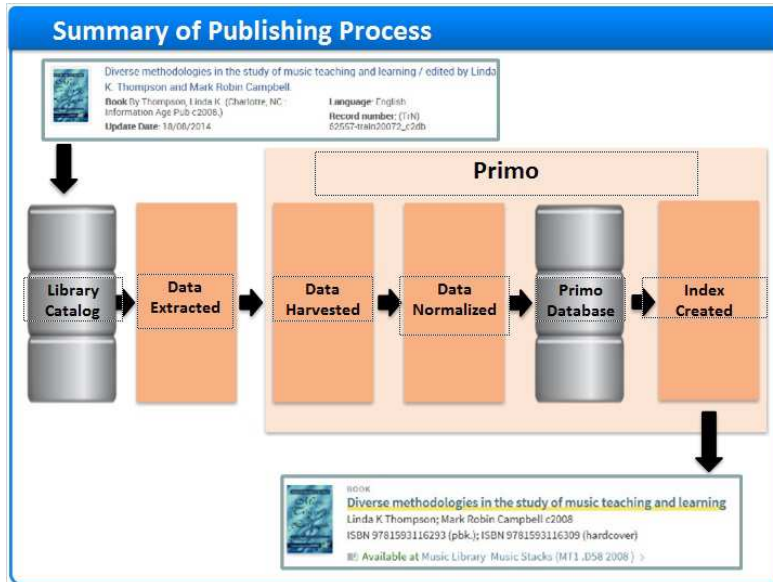
Notes:

From the original record in the source system to the search result in Primo, the publishing process ensures that your users will be able to search across all of your various data sources and retrieve results that are consistently organized and displayed.

1.19 Session Review

Notes:

Session Review

1.20 Session Review

Notes:

In this session we covered how your data gets loaded into Primo, what happens to your records when they are loaded, what these new records look like.
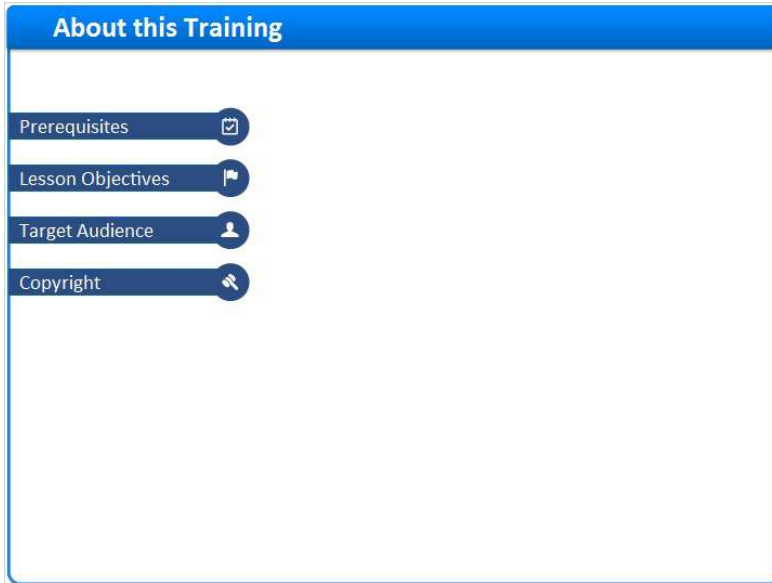
1.21 Thank You!

Notes:

Thank you for joining this session!

1.22 About this Training

Notes:

1 (Slide Layer)

2 (Slide Layer)

About this Training

Prerequisites

Lesson Objectives

Target Audience

Copyright

By the end of this training session, you will know:

- How your data gets loaded into Primo
- What happens to your records when they are loaded
- What these new records look like

3 (Slide Layer)

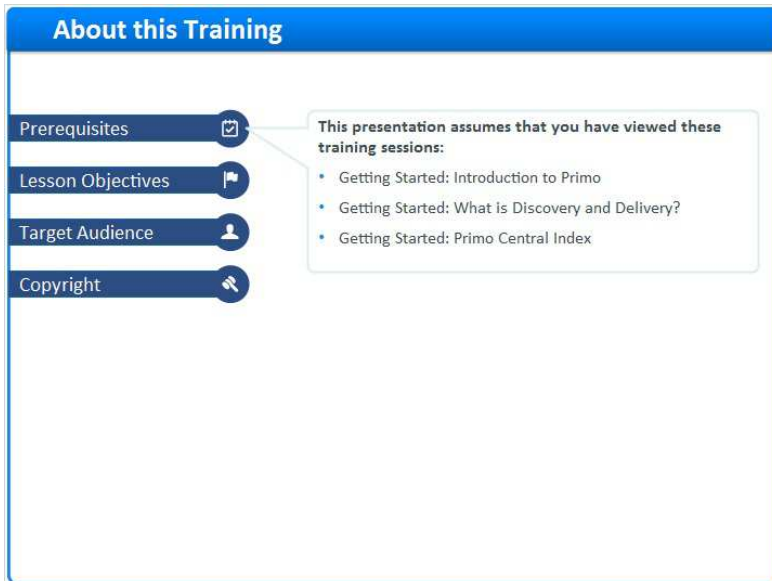**About this Training**

Prerequisites

Lesson Objectives

Target Audience

Copyright

Target Audience:

- Staff maintaining Primo

4 (Slide Layer)

About this Training

Prerequisites

Lesson Objectives

Target Audience

Copyright

All of the information and material inclusive of text, images, logos, product names is either the property of, or used with permission by Ex Libris Ltd. The information may not be distributed, modified, displayed, reproduced – in whole or in part – without the prior written permission of Ex Libris Ltd.

TRADEMARKS
Ex Libris, the Ex Libris logo, Aleph, Alma, SFX, SFXIT, MetaLib, DigiTool, Verde, Primo, Voyager, MetaSearch, MetaIndex, Rosetta, Leganto and other Ex Libris products and services referenced herein are trademarks of Ex Libris, and may be registered in certain jurisdictions. All other product names, company names, marks and logos referenced may be trademarks of their respective owners.

DISCLAIMER
The information contained in this document is compiled from various sources and provided on an "AS IS" basis for general information purposes only without any representations, conditions or warranties whether express or implied, including any implied warranties of satisfactory quality, completeness, accuracy or fitness for a particular purpose.

Ex Libris, its subsidiaries and related corporations ("Ex Libris Group") disclaim any and all liability for all use of this information, including losses, damages, claims or expenses any person may incur as a result of the use of this information, even if advised of the possibility of such loss or damage.

© Ex Libris Ltd., 2018