ESPLORO ESSENTIALS: *Smart Harvesting*

Automated Repository Population

How do you build a comprehensive repository and keep it updated as your researchers continue to publish?

In a previous training video, we demonstrated how to manually add individual Research Assets to your repository.

In this video, we will describe three ways Esploro can automate repository population:

Smart Migration, Smart Expansion, and Smart Harvesting.

Smart Migration takes metadata from your old institutional repository and automatically creates entities and relationships, enriching them so that they fit into the Esploro data model. For example, the authors from each publication are matched with researchers in your Esploro system.

Smart Expansion starts with publications known to be authored by an affiliated researcher. These are found using identifiers, such as researcher IDs and DOIs, or using citation lists. Smart Expansion creates an asset record for each publication and matches all the co-authors to Esploro researchers.

Smart Harvesting, on the other hand, is exploratory. Instead of retrieving records for known publications, it searches the internet for assets relevant to your institution's researchers.

Smart Harvesting leverages CDI, Ex Libris's Central Discovery Index. CDI has billions of records from multiple sources and ingests more daily, so it is a great way to search many publication platforms.

Smart Harvesting uses the researcher name to find relevant assets for that researcher. Most records do not contain author IDs like ORCID, or author email addresses, which could unambiguously connect the researcher to the asset .

Matching author names to researchers is not a simple process. A researcher may have published under multiple different names, and those names may appear in multiple formats. Also, names are not unique; different researchers can have the same name, even in the same field of research.

To overcome these obstacles and determine the likelihood that an asset was indeed written by a particular researcher, Smart Harvesting uses an author matching algorithm developed using Machine Learning methodologies.

Smart Harvesting gathers candidate assets and ranks the strength of each author match. Your repository manager controls which author matches require approval by administrators and which will be approved automatically, as we will see in the next video. Assets with approved author matches become part of your institution's repository.

The author matching algorithm relies on researcher profile information and assets already associated with the researcher in Esploro, so the more information there is – the better. Therefore, Smart Migration and Smart Expansion should be run before Smart Harvesting.

All of these methods enable you to enrich your repository more efficiently.