Reimagining Course Discovery

Open edX Conference 2023

Agenda

- A 38000ft overview of Discovery

- Can we expand the business use cases for discovery?

    - Can discovery act as an EdTech marketing platform?

Speakers

Proposed & Prepared

Presenters

Syed M. Dawoud Sheraz Ali

Syed Ansab Waqar Gillani

Irfan Ahmad

Umar Ali Choudhry

Course Discovery

- Course Catalog - designed for Open edX ecosystem

- Consolidated source of metadata for different products in the Open edX ecosystem

    - Powers search data on edx.org

    - Catalog APIs for external parties

- Pushes data to various Open edX services such as Studio & Ecommerce

Course Discovery

- Important Components

  - Courses (including Course Run)

  - Program & Degree

  - Partner Organizations

  - Course Editors

  - Search focused information

Course Discovery

- Data loaders for ingesting data across services (Studio, Ecommerce, Analytics)

- Data Indices capabilities for search & data filtering

  - ElasticSearch – Catalogs

  - Algolia – edx.org

Pathways for your advancement

Compare Programs

Executive Education `New`
Expert-led, fully supported courses that build career-critical skills

Professional Certificates
From employers or universities to build today's in-demand skills

Master's Degrees `New`
Online degree programs from top universities

MicroMasters® Programs
Graduate-level, for career advancement or a degree path

Bachelor's Degrees `New`
Begin or complete a degree; fully online

MicroBachelors® Programs
Undergraduate-level, for career advancement or a degree path

XSeries
Series of courses for a deep understanding of a topic

Gain Career-Critical Skills
Courses built to fit your busy lifestyle and help move your career forward.

Group-Based Learning
Engage with subject matter experts and network with fellow working professionals.

24/7 Support
Around the clock technical assistance, along with support from expert tutors and dedicated advisers.

Explore Executive Education Courses `New`

AI & Digital Transformation

Artificial Intelligence: Implications for Business Strategy
MIT Sloan School of Management

Executive Education

Oxford Artificial Intelligence Programme
Oxford Saïd

Executive Education

Machine Learning in Business
MIT Sloan School of Management

Executive Education

Automation: Implementation in Business
LSE

Executive Education

Show 4 more in AI & Digital Transformation

Explore bachelor's degrees `New`

Business

Bachelor of Science in Business and Management
University of London

⧉ Bachelor's Degree

Bachelor of Science in International Relations
University of London

⧉ Bachelor's Degree

Bachelor of Science in Accounting and Finance
University of London

⧉ Bachelor's Degree

Bachelor of Science in Economics and Finance
University of London
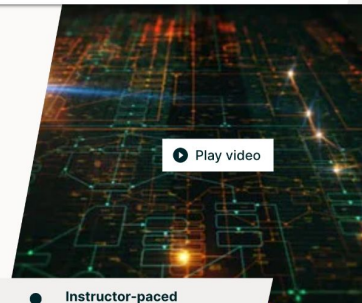
⧉ Bachelor's Degree

MIT

Machine Learning in Business

Incorporate machine learning into your business strategy, as you explore the value and impact of this technology.

This course is offered by GetSmarter, an edX partner

Request more info

▶ Play video

◔ **6 weeks**
6–8 hours per week

📅 **Starts May 3, 2023**
Enroll by Apr 25, 2023

👤 **Instructor-paced**
Instructor-led on a course schedule

Request a **course brochure**

Want to know more? Enter your information to learn more about this course from GetSmarter (an edX partner).

First Name

Last Name

Email

☐ Email me about the MIT Machine Learning in Business online short course

By consenting to receive communications, you agree to the use of your data as described in GetSmarter's privacy policy. You may opt out of receiving communications at any time.

Request more info

Why choose GetSmarter?

getsmarter

GetSmarter, an edX partner, delivers career-relevant, fully supported, and expert-led online courses from the world's leading universities and institutions. As an online learning expert with over 15 years' experience, GetSmarter leverages long-standing expertise in market research, data analysis, and business insights to ensure that you gain the skills critical to career growth.

Read more about GetSmarter ↗

New products!

- New products on edx.org
- No enrollment option
- edX partner – GetSmarter, for Exec Ed

External Products in Discovery

- Since 2U acquisition, new products have been added on edx.org
- Products are on edx.org but consumed off-site (GetSmarter, for instance)
- edx.org is powered by Discovery!

External Products in Discovery

- Capability in Discovery to market or showcase products external to platform
- Limited by core design of Discovery platform centered around edx platform

... Eureka! ... convert this challenge into an opportunity!

Challenges

- Cyclic data synchronization between Discovery, Studio & Ecommerce
- Multiple sources of truth
  - Schedule dates: required by Discovery on creation but only updateable in Studio
- Development
  - ElasticSearch as a hard requirement

Course Publishing Flow - Data Synchronization Timings
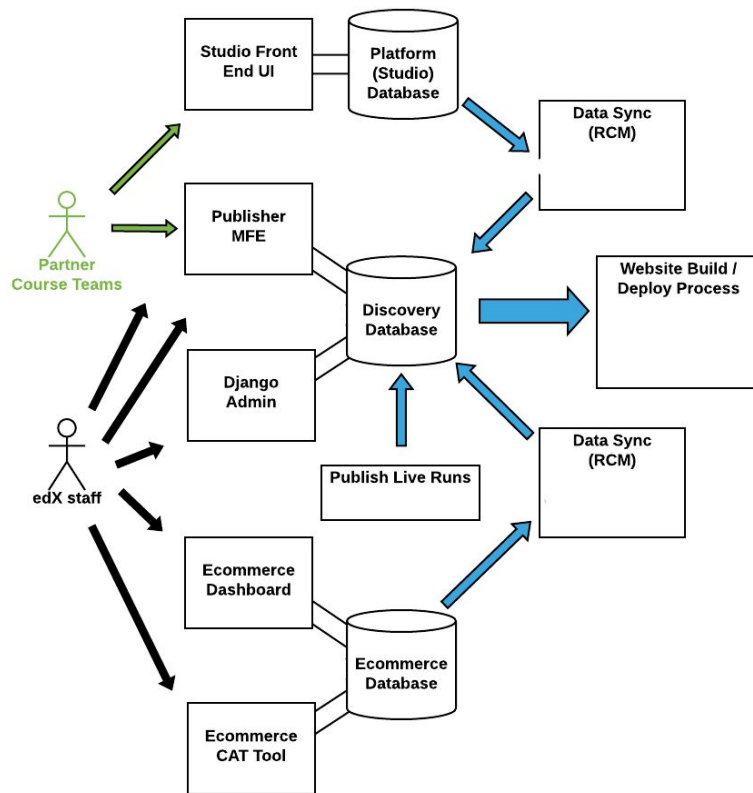
Challenges

- Course & Course run types
  - setup needed in Discovery, Studio/LMS, and E-commerce to add new types
- Draft Mixins
  - Maintaining 2 versions of objects like Course, CourseRun
  - Data tagged from draft → non-draft under certain conditions as part of APIs

Reimagining

A marketing platform for EdTech space capable of integrating with Open edX

Reimagining

- Support the existing Open edX centered design and enhance to accommodate external products

- … friction free integration!

The Integration Challenge

Single Source of Truth vs Multiple Sources of Truth

Partial conflict resolution using data partition by source

Alternates?

Impacted Areas

1. Data Models
2. Data Ingestion
3. Data Index
4. Open edX connectivity
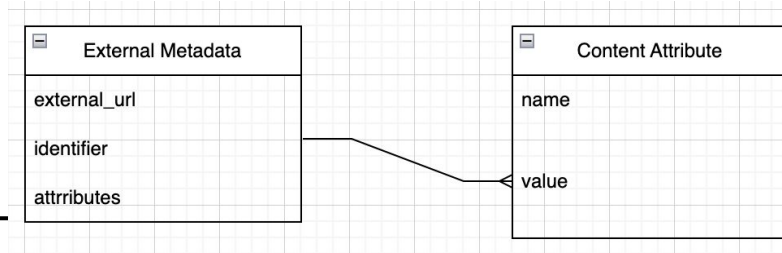
Product Data Model Changes

- Partition the current metadata
- Lets partition!
  - Categorizing data fields
    - Internal Metadata
    - External Metadata
    - Search Metadata
    - Product Classification
    - Line of Business classification
    - Product Source

Data Model Changes - Internal Metadata

- A Mix of required and internal fields
  - Title, description, subject, learning outcomes, cost/price
  - Introductory Video, Instructors
- These are fields that will help learners decide if they should be taking that product
- Give more control to editors & coordinators to manually flip the product statuses to in product pipeline

Data Model Changes - External Metadata

- Required data for ingesting external products
- Can include additional information not present in Internal Metadata
  - Dynamically set in form of list of {"attr": "value"}
- Base external metadata information
  - Redirect URL
  - External system product identifier

Data Model Changes - Search Metadata

- Collection of fields or data models used solely for search purposes
- Generic + Specialized fields
  - Product geolocation
  - Search rank value
  - Product lead value
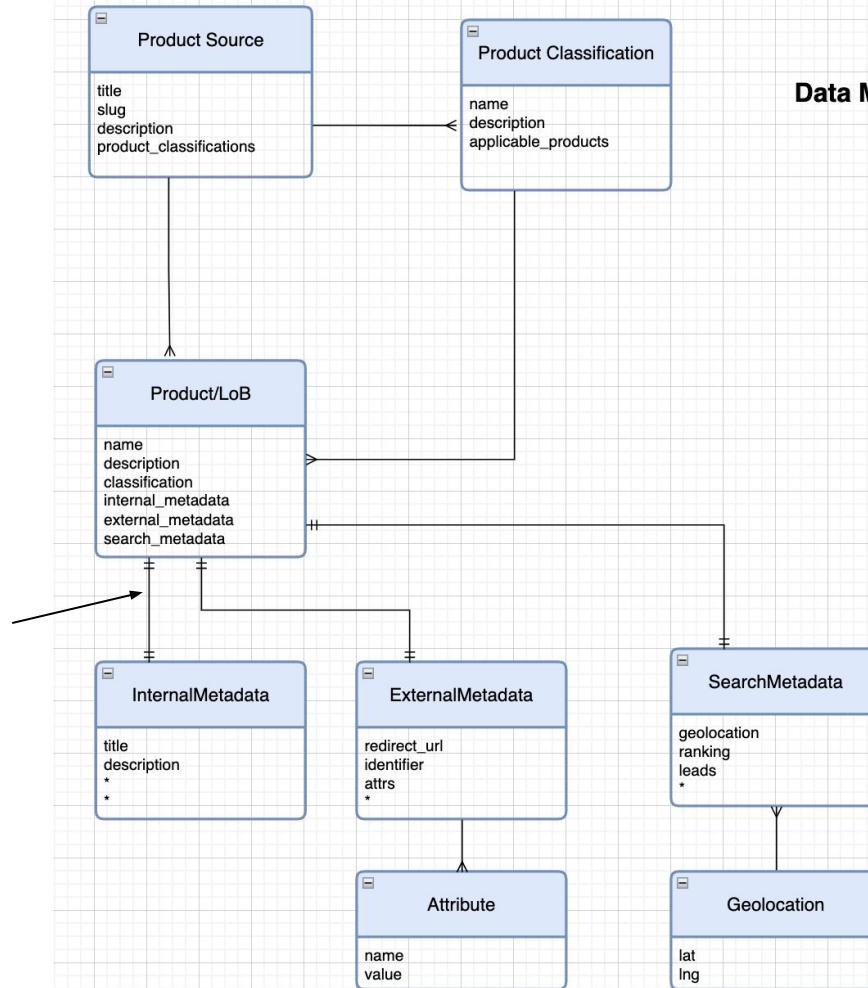 - Should appear in search results?

Data Model Changes - Product Source

- Information identifying whether a product is from the platform or from an external source
- Influences the processing of the product in the edx platform

Data Model Changes - Product Line & Classification

- Similar to current Course & Program types design
- Product → Product Line (ExecEd, Degree, Course etc.)
- Product Type → Product Classification
- Examples
  - Line → ExecEd, Classification → MicroCourse
  - Line → Course, Classification → Verified
  - Line → Degree, Classification → Bachelors

Data Model -- Abstract

Product Source

title
slug
description
product_classifications

Product Classification

name
description
applicable_products

Product/LoB

name
description
classification
internal_metadata
external_metadata
search_metadata

Could be Proxy models also, depending upon implementation

InternalMetadata

title
description
*
*

ExternalMetadata

redirect_url
identifier
attrs
*

SearchMetadata

geolocation
ranking
leads
*

Attribute

name
value

Geolocation

lat
lng

Impacted Areas

1. Data Models
2. Data Ingestion
3. Data Index
4. Open edX connectivity

Data Ingestion

- Bulk Ingestion
  - CSV Data loaders for bulk import of data, both internal and external products
  - Generic interface to get bulk data into the system
- Unified API
  - Single entry point for internal/external parties to add/update products
  - Event driven, bus friendly, design

Data Index

- Plugins or services will use Discovery API to index data in Index service of choice
  - ElasticSearch
  - OpenSearch
  - Algolia
- Flexibility in defining the depth and breadth of indexed data

Open edX Connectivity

- Control data push to services based on Product source
- Configure external data push in Django models
  - External resource url
  - Request body & method
- Dynamically set request data in as list of {"attr": "value_type"} using Django models
- Use OAuth to authenticate (configured on Product source level)
- Similar to existing Partner model in Discovery
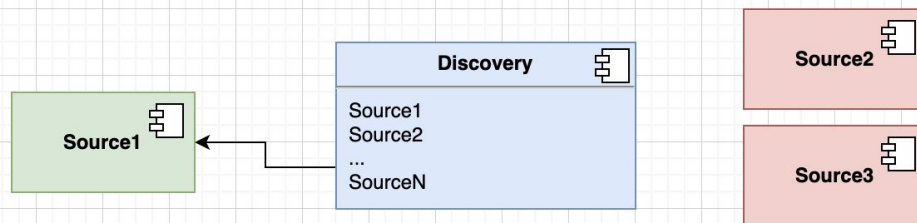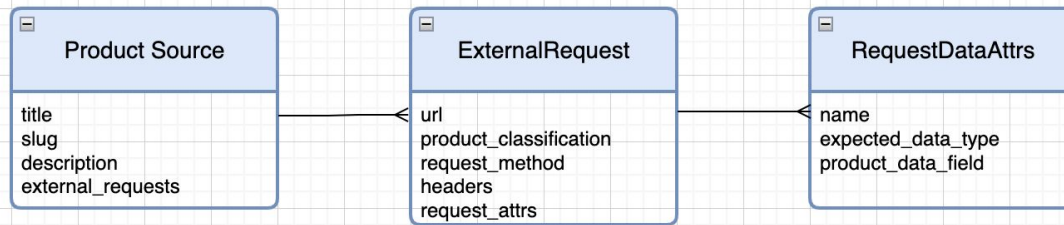
Connecting it all together

- The Open edX centric design of Discovery made it challenging to seamlessly represent products delivered from external sites.
- Great opportunity to align Discovery as a product marketplace
- *Eventually Consistent* Multiple Sources of Truth

- Will the community adopt Discovery as a marketing platform? … we think so!
  - Regional Open edX deployments marketing content from lesser known platforms!

References

- https://edx-discovery.readthedocs.io/en/latest/introduction.html
- https://edx.readthedocs.io/projects/edx-partner-course-staff/en/latest/set_up_course/pub_create_ann_course/pub_introduction.html
- https://discuss.openedx.org/t/when-to-make-a-new-backend-service/8267
- https://django-model-utils.readthedocs.io/en/latest/utilities.html#field-tracker
- https://github.com/openedx/course-discovery/blob/master/docs/decisions/0004-publisher-draft.rst

[Discuss.openedx post](https://discuss.openedx.org)

When to make a new backend service? ✏️

Development Architecture

dave 🛡️ David Ormsbee   Sep '22

There's recently been talk about when to make something a service vs. making separate Django apps that plug into the LMS process at runtime. I thought there was specific guidance on this in the wiki somewhere, but I couldn't find it. So I'm curious to get people's thoughts.

Someone recently framed it as a neat thought experiment: **If we were building course-discovery 1 today, would it still be a separate service?** Or would we prefer to make a separate set of apps that runs in-process in the LMS?

It was originally built as a separate service for a number of reasons:

1. **Faster Deployment Cycles**
   At the time that course-discovery was created, the deployment cycle for the LMS was somewhere between one and two weeks. Missing the release window was extremely painful. In contrast, a new service that didn't carry all of edx-platform's deployment baggage with it could run through CI/CD and have new commits be in production in under an hour.
2. **Team Autonomy**
   Having a shared service meant that you always had to worry about being delayed by another team—someone causing a bug that forces a rollback, or another team asking for a short delay in the release so they could their one critical feature over the line and not have to wait until the following week. Having a separate service meant that your team was fully in control.
3. **Operational Independence**
   In theory, having separate services gives a certain amount of resiliency, since a failure of one component doesn't take down everything else with it. In practice, this is a mixed bag because many synchronous API dependency calls do exist between the LMS and other services, meaning that failures cascade anyway. In addition, the added data access challenges means that some of these calls are much more expensive than they might otherwise be (e.g. n+1 queries via REST API instead of doing a SQL join).
4. **Force Separation of Logic**
   If you're in the same process, it's much easier to throw in a quick hack and import some internal piece you're really not supposed to. The thought here was that physical boundaries forced developers to think about these interfaces in a more rigorous way.

Looking at the discovery service documentation, here is its explanation:

> The distribution of edX's data has grown over time. Any given feature on edx.org may need information from Studio, the LMS, the Ecommerce service, and/or the Drupal marketing site. Discovery is a data aggregator whose job is to collect, consolidate, and provide access to information from these services.

At one level, that seems very reasonable. And if the data flow were like this:

I think it would be great.

But as I understand it, it's really more like this in practice: