



Development of a Versatile, Full-Featured Search Functionality for Indico

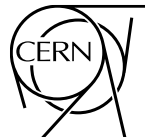
Penelope Constanta

CHEP 2019

4 November 2019

In collaboration with:

BROOKHAVEN
NATIONAL LABORATORY



The Collaboration

- **Fermilab**
 - Penelope Constanta
- **BNL**
 - Ofer Rind
 - Jose Caballero Bejar
- **CERN**
 - Pedro Ferreira
 - Adrian Mönnich
 - Pablo Panero
 - Carina Rafaela De Oliveira Antunes
 - Aristofanis Chionis Koufakos

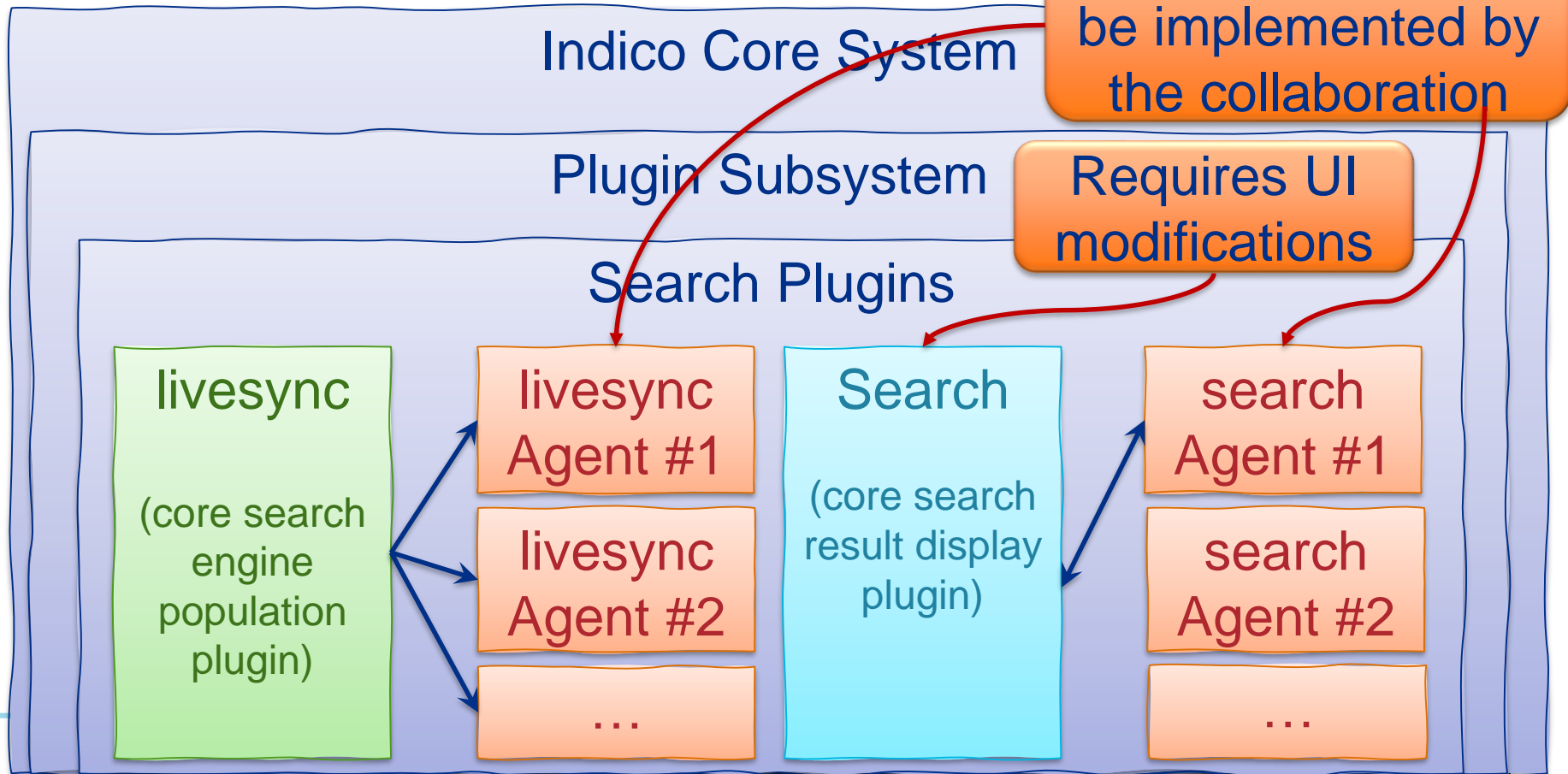
Overview

- **Indico is:**
 - an open-source event management system, popular in HEP community
 - extensible through its plugin architecture (PayPal, video conferencing, search etc.)
- **Indico v2.x:**
 - has many improvements throughout the system
 - lacks search capabilities, outside the CERN eco-system that uses SharePoint
- **Search plugin necessity:**
 - CERN is moving away from SharePoint by the end of this year, to the new invenio based *CERN Search μ service*, necessitating the development of an indico interface
 - Fermi and BNL user communities requested a full functional search before deploying the new indico version
- **Fermi-BNL-CERN collaboration to build the search plugins:**
 - Utilizing the new *CERN Search μ Service* and make it available to the community

Indico Search

- **Indico v0.98 – v1.2:**
 - search utilizes invenio (v1.1) as its search engine sending its metadata in XML format
 - search results are formatted and displayed appropriately by indico
 - Framework can be used outside CERN's environment
- **Indico v1.9 – v2.2:**
 - search sends search metadata to SharePoint by re-purposing the existing invenio plugin code
 - Metadata formatting does not take advantage of the new python packages (SQLAlchemy, marshmallow, etc.)
 - search results are displayed by SharePoint (indico simply displays the SharePoint page)
 - Framework cannot be used outside CERN's environment
- **Collaboration plugin development for next version of indico v2.2.x:**
 - search utilizes invenio's (v3) *CERN Search Api* component and Elasticsearch as its search engine, sending its metadata in JSON format and taking advantage of SQLAlchemy and marshmallow
 - search results are formatted and displayed appropriately by indico
 - framework is developed so that it can be used outside CERN's environment

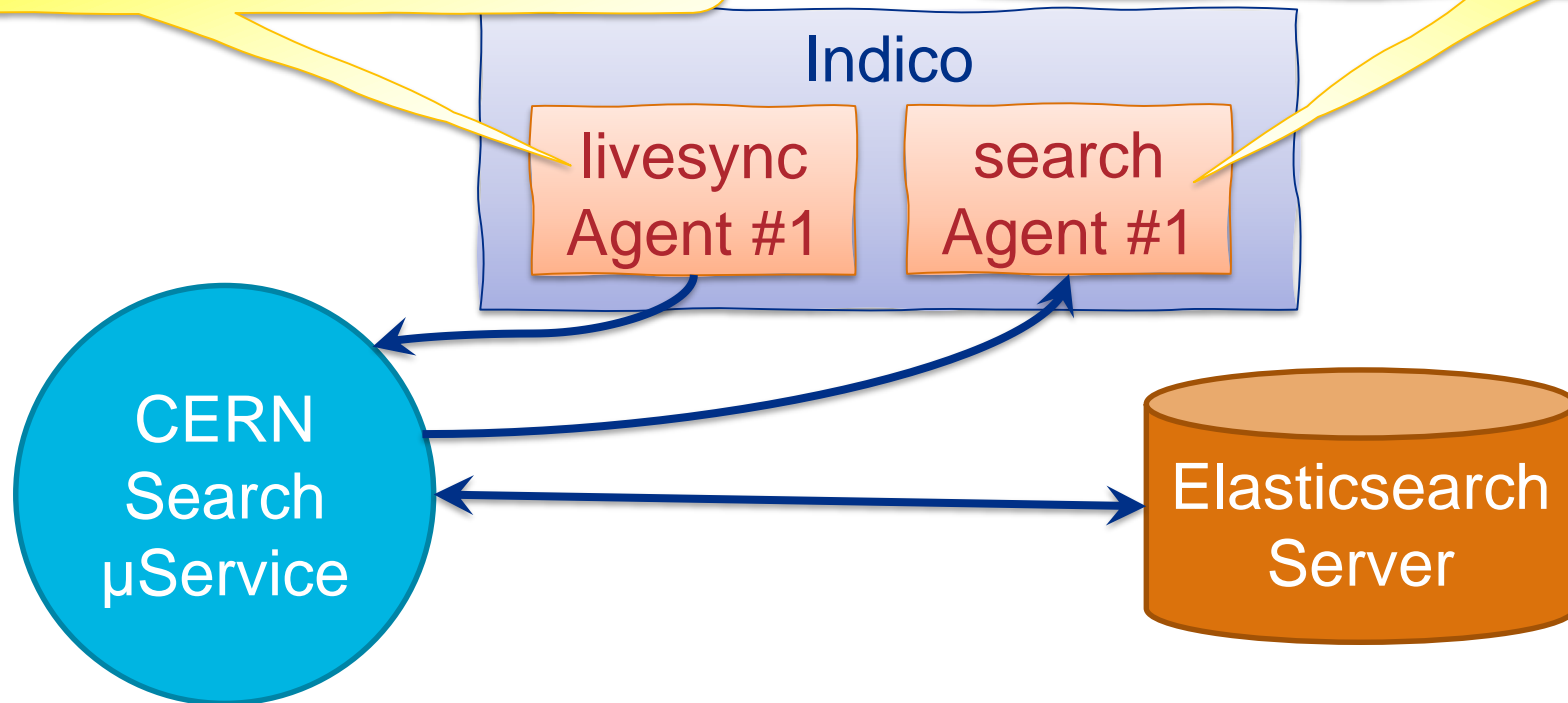
Indico Code Architecture



Indico Search System Architecture

Sends indico object metadata to https://cernsearch_api/

Agent that sends search strings, receives & displays search results (https://cernsearch_api/)



Implementation Challenges

- Indico v2.x moved away from the ZOPE database to PostgreSQL and almost the entire indico code was re-written and restructured
 - Any familiarity with the previous versions' code is not useful
 - Plugin development is seemingly easier but at the end one needs to understand all the internals of the new indico plugin system as well as the interface with the base plugins and the core indico code, along with the numerous new python packages
- CERN Search μ service is very new and documentation is targeted for CERN's internal use
 - Deployment through docker-compose prove to be more challenging as the μ service is targeted for CERN's internal use.
- FNAL and BNL developers worked for a fraction of their time on the indico project and were not familiar with the used python packages.

Indico 2.x Installation / Configuration

- CERN's documentation is **excellent** for installing/upgrading and setting up indico!
- Installation
 - Just follow CERN's indico 2.x installation
 - <https://docs.getindico.io/en/latest/installation/production/>
 - For our development purposes we installed the developer's version:
 - <https://docs.getindico.io/en/latest/installation/development/>
- Enable search plugins – Configuration
 - All required steps are at:
 - <https://docs.getindico.io/en/latest/installation/plugins/>

Deployment of CERN Search μ service

- CERN provided the docker-compose.yml that creates:
 - The cern_search_rest_api as an invenio component
 - The PostgreSQL application
 - not required if connecting to an existing DB
 - The Elasticsearch (ES) application
 - not required if connecting to an existing ES installation
 - The ES kibana application
 - not required if connecting to an existing ES installation
 - The tika server to parse PDF, pptx, LaTeX etc. files, needs to be added, if not connecting to an existing tika server.
 - It also initializes the invenio DB and uploads the ES mappings

Implementation Status

- livenessync agent for CERN Search μ service
 - First version development almost completed
 - Not fully tested, awaiting the cern-search-api deployment
- Indico search User Interface
 - First version development completed, requires minor modifications
 - Search results UI:
 - provides filtering capabilities for Speakers and Affiliations
 - uses different tabs for events, contributions, attachments, notes
 - displayed page controls
 - Tested with mock data
- search agent for CERN Search μ service
 - Last stages of development
 - Not fully tested, awaiting the cern-search-api deployment

Indico livenessync Agent plugin Configuration

LiveSync_Json is the agent for the CERN search μ service

The screenshot shows the Indico web interface. The top navigation bar includes 'Home', 'Create event', 'Room booking', 'Administration', and 'My profile'. Below this, a breadcrumb trail reads 'Home > Administration > Plugins'. A sidebar menu on the left contains 'General Settings', 'Security', 'Plugins' (highlighted), 'Room Booking', 'User Management', 'Customization', 'Homepage', 'Integration', and 'Tasks'. The main content area is titled 'Plugins' and 'Synchronization', featuring two buttons: 'LiveSync' (version 2.0) and 'LiveSync_JSON' (version 1.0).

The 'LiveSync Plugin Settings' page provides basic LiveSync functionality. It includes a 'Queue entry TTL' field set to '0', with a description: 'How many days should processed entries be kept in the queue. Time counts from the creation of the queue entries, so if the Live task is not running for some time, queue entries may be deleted during the next run after processing them. Setting it to 0 disables automatic deletion.' Below this is an 'Excluded categories' section with a 'Category ID' input field and an 'Add' button. A note states: 'Changes to objects inside these categories or any of their subcategories are excluded.' At the bottom, there are 'Save' and 'Cancel' buttons. The 'LiveSync Agents' section shows a message: 'No agents have been added yet.' and an 'Add LiveSync-JSON agent' button.

The 'LiveSync_JSON Plugin Settings' page provides the LiveSync-JSON agent for LiveSync. It contains several configuration fields: 'Search app URL' (with description: 'URL of search app import endpoint'), 'Search app TOKEN' (with description: 'TOKEN for accessing the Search app import endpoint'), 'Elasticsearch Events JSON Schema' (set to 'events_v1.1.0.json', description: 'the JSON Schema for the events Elasticsearch index'), 'Elasticsearch Contributions JSON Schema' (set to 'contributions_v1.1.0.json', description: 'the JSON Schema for the contributions Elasticsearch index'), 'Elasticsearch SubContributions JSON Schema' (set to 'subcontributions_v1.1.0.json', description: 'the JSON Schema for the subcontributions Elasticsearch index'), 'Elasticsearch Attachments JSON Schema' (set to 'attachments_v1.1.0.json', description: 'the JSON Schema for the attachments Elasticsearch index'), 'Elasticsearch Notes JSON Schema' (set to 'notes_v1.1.0.json', description: 'the JSON Schema for the notes Elasticsearch index'), and 'tika server URL' (with description: 'URL of tika server to parse file content. If not supplied a local tika server will be instantiated.'). 'Save' and 'Cancel' buttons are located at the bottom.

Indico Search User Interface

The main navigation bar features the Indico logo on the left. On the right, it includes a status bar with 'Public', 'Europe/Zurich', and 'P. Ferreira'. Below this is a dark navigation menu with links for Home, Create event, Room booking, Administration, Services, and My profile. A search bar is positioned above the main content area, containing the text 'CHEP2019' and a 'Create event' button.

Welcome to Indico. The Indico tool allows you to manage complex conferences, workshops and meetings. To start browsing, please select a category below.

- CERN-Related
- Committees
- Conferences, Workshops and Events
- Departments
- Directorate

This panel shows a search for 'CHEP 2019'. On the right, a list of event counts is displayed: 6,226 events, 7,367 events, 7,574 events, 60,202 events, and 234 events. On the left, a sidebar lists 'Speakers' (Aristofanis, Pedro, Adrian) and 'Affiliations' (Natalia's, Marco's, Giota's). The main content area shows search results for 'CHEP 2019', including 'CHEP 2019 Full PC' and 'CHEP 2019 IAC'.

This panel shows a search for 'enterprise'. The search results are filtered to show 6 events. The first result is 'Cells-as-a-Service: Enterprise-Grade Cloud Infrastructure Research at HP Laboratories'. Other results include 'Integrating Oracle VM into an Enterprise-Grade OpenStack Cloud: CERN Case Study', 'Manage the Manager: Tips on How to Best Manage Oracle Enterprise Manager 12c (CONB741)', 'Massive Predictive Modeling', 'Systems Engineering Workshop', and 'Virtual Visit : Norbury Manor Business and Enterprise College for Girls'.

A callout box highlights a search result for 'Indico CONFERENCE: Candidate participant's registration/application' dated 18 July 2017. The result is categorized under 'Schools, Seminars and Courses > Training and Development'. A pagination bar at the bottom shows page 1 of 8.

Future Development

- All plugins developed by the collaboration will be integrated into indico and CERN will take ownership.
- Further development may include:
 - Improved resilience and recovery for the livesync agent
 - Extensions to search UI, if needed
 - Improved developer documentation and deployment for non-CERN environments

If you can find this talk on CERN's indico site, using indico search, in 2020 then this collaboration was successful!