



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

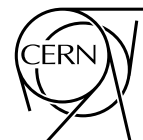
Penelope Constanta

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

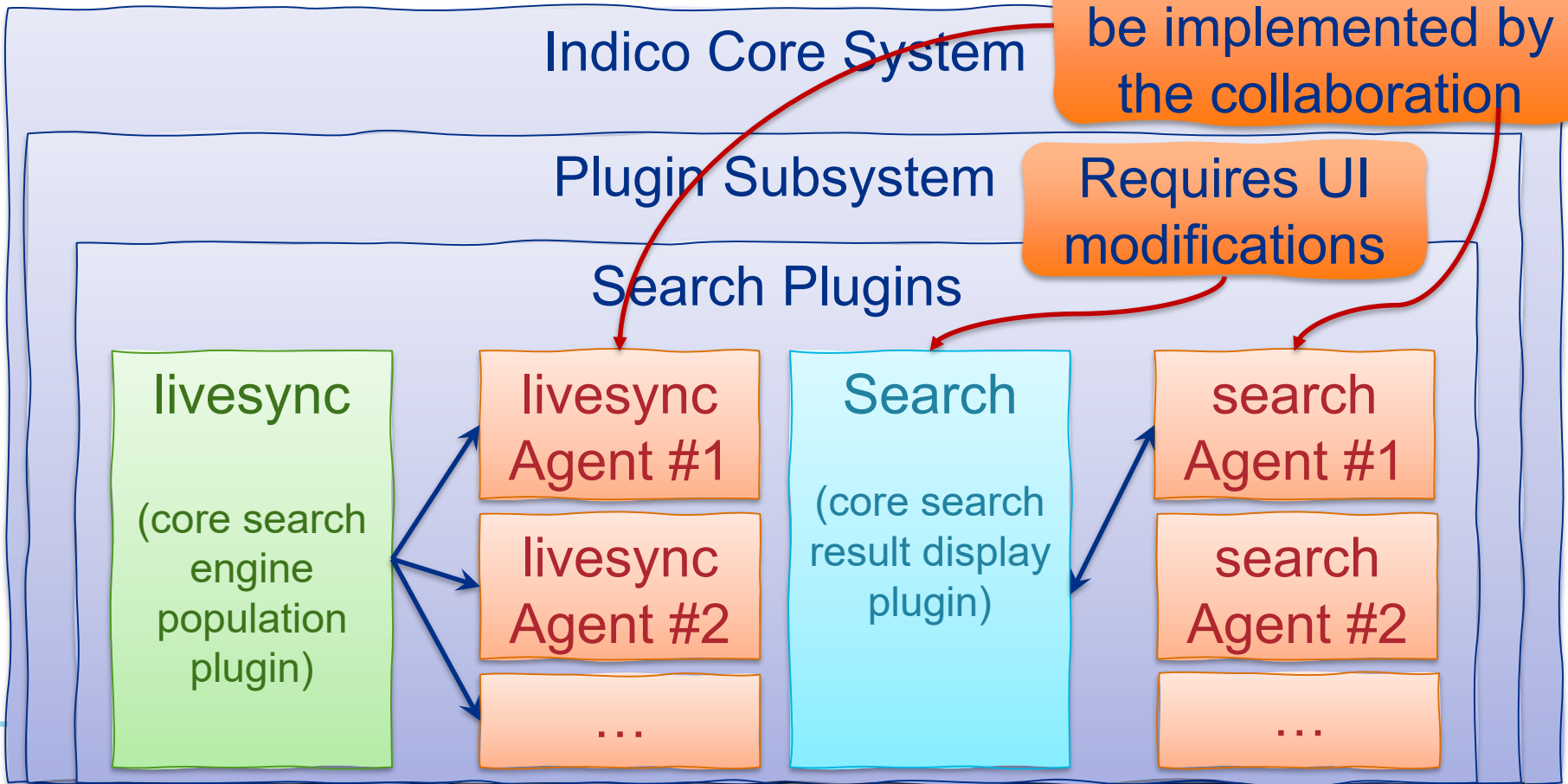
# Overview

- **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  $\mu$ 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  $\mu$ 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

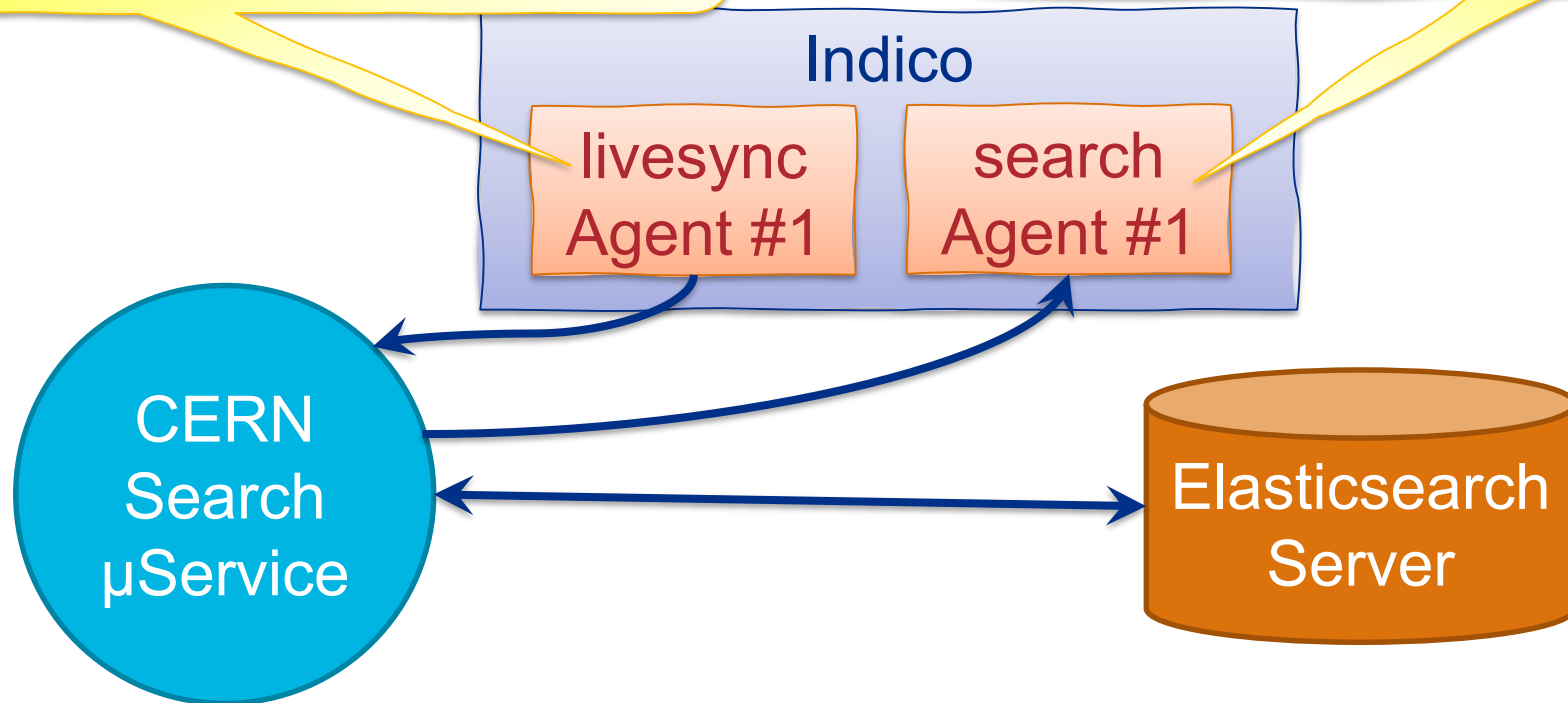
# Indico Code Architecture



# Indico Search System Architecture

Sends indico object metadata to [https://cernsearch\\_api/](https://cernsearch_api/)

Agent that sends search strings, receives & displays search results ([https://cernsearch\\_api/](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  $\mu$ service is very new and documentation is targeted for CERN's internal use
  - Deployment through docker-compose prove to be more challenging as the  $\mu$ 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 $\mu$ 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  $\mu$ 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  $\mu$ 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  $\mu$ service

The screenshot shows the Indico Administration interface. The top navigation bar includes 'Home', 'Create event', 'Room booking', 'Administration', and 'My profile'. The breadcrumb trail is 'Home > Administration > Plugins'. A sidebar on the left lists various settings: General Settings, Security, Plugins (selected), Room Booking, User Management, Customization, Homepage, Integration, and Tasks. The main content area is titled 'Plugins' and 'Synchronization', showing two active plugins: 'LiveSync' (version 2.0) and 'LiveSync\_JSON' (version 1.0).

The 'LiveSync Plugin Settings' window provides basic configuration. 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 'Save' button is at the bottom. 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' window provides configuration for the JSON agent. It includes several fields: 'Search app URL' (URL of search app import endpoint), 'Search app TOKEN' (TOKEN for accessing the Search app import endpoint), 'Elasticsearch Events JSON Schema' (events\_v1.1.0.json), 'Elasticsearch Contributions JSON Schema' (contributions\_v1.1.0.json), 'Elasticsearch SubContributions JSON Schema' (subcontributions\_v1.1.0.json), 'Elasticsearch Attachments JSON Schema' (attachments\_v1.1.0.json), 'Elasticsearch Notes JSON Schema' (notes\_v1.1.0.json), and 'tika server URL' (URL of tika server to parse file content). 'Save' and 'Cancel' buttons are at the bottom.

# Indico Search User Interface

Indico logo | Public | Europe/Zurich | P. Ferreira

Home | Create event | Room booking | Administration | Services | My profile

Main categories | Search: CHEP2019 | Create event

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

Search: CHEP 2019

Categories (5) | Events (0) | Contributions (0) | Materials (0)

**Speakers**

- Aristofanis (45)
- Pedro (4555)
- Adrian (455)

**Affiliations**

- Natalia's (123)
- Marco's (321)
- Giota's (333)

**CHEP 2019**  
Conferences, Workshops and Events > Conference Organisation > CHEP

**CHEP 2019 Full PC**  
Conferences, Workshops and Events > Conference Organisation > CHEP > CHEP 2019

**CHEP 2019 IAC**  
Conferences, Workshops and Events > Conference Organisation > CHEP > CHEP 2019

**CHEP 2019 LOC**  
Conferences, Workshops and Events > Conference Organisation > CHEP > CHEP 2019

6,226 events

7,367 events

7,574 events

60,202 events

234 events

Search

enterprise

Categories (0) | **Events (6)** | Contributions (1) | Materials (1)

**Speakers**

- Aristofanis (45)
- Pedro (4555)
- Adrian (455)

**Affiliations**

- Natalia's (123)
- Marco's (321)
- Giota's (333)

**Cells-as-a-Service: Enterprise-Grade Cloud Infrastructure Research at HP Laboratories**  
24 March 2019 14:00  
Schools, Seminars and Courses > Seminars

**Integrating Oracle VM into an Enterprise-Grade OpenStack Cloud: CERN Case Study**  
01 October 2014 12:45  
Departments > IT > Groups > DB

**Manage the Manager: Tips on How to Best Manage Oracle Enterprise Manager 12c (OCN8741)**  
02 October 2014 22:15  
Departments > IT > Groups > DB

**Massive Predictive Modeling**  
08 April 2014 11:00  
Schools, Seminars and Courses

**Systems Engineering Workshop**  
15 October 2010 10:00  
Projects > MedAustron

**Virtual Visit : Norbury Manor Business and Enterprise College for Girls**  
05 Jul 2016 18:30

Indico CONFERENCE: Candidate participant's registration/application  
18 July 2017 18:00  
Schools, Seminars and Courses > Training and Development

1 2 3 4 5 6 8

Help | Contact | Terms and conditions

## 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!**