Radiam

Scalable Metadata Indexing For Distributed Research Data

Todd Trann
Technical Lead

Canadian Research Software Conference
May 29, 2019

University of Saskatchewan | SFU Simon Fraser University Engaging the World | compute canada | calcul canada | canarie

WHAT IS RADIAM?

- Radiam is a centralized searchable metadata index for the distributed data of a research project during active data collection and processing
- Built on top of an existing open source, scalable search engine (Elasticsearch)
- Offers up a secure API that can be used by various plugins and agents to update and query the index
- Enables rich, standards-based metadata application near the point of data collection
- Read more: https://www.radiam.ca

PROJECT BACKGROUND

- The University of Saskatchewan and Simon Fraser University, with the support of Compute Canada/WestGrid and CARL/Portage proposed the project in spring of 2018
- CANARIE’s Research Software Program provides 18 months of project funding (Oct 2018 – Mar 2020)
- On completion, Radiam will be open source and licensed without restrictions so that it can be used for future projects

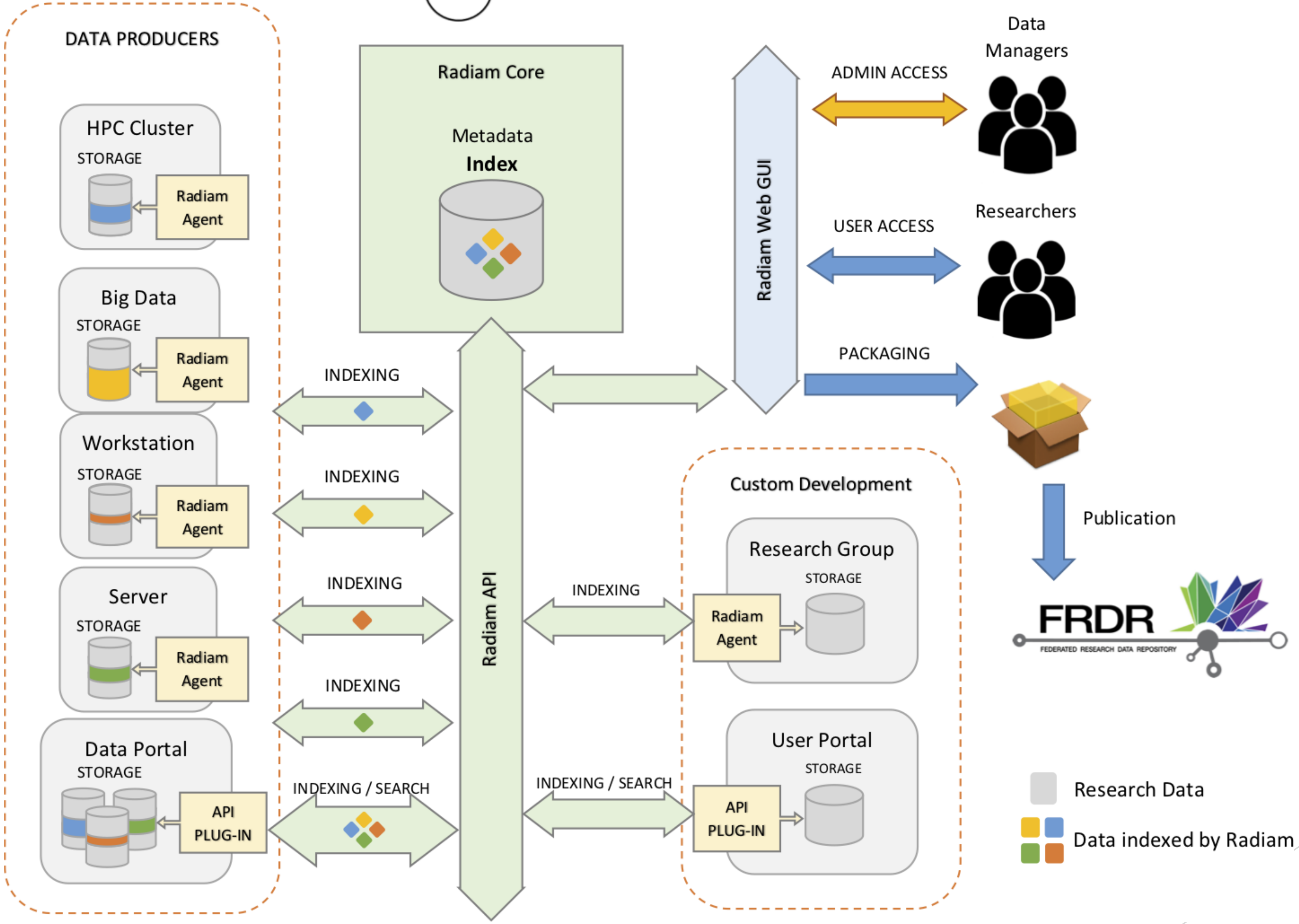
FEATURES

- **Index**: crawl locations for data, submit metadata to the search index
- **Annotate**: augment collected metadata with additional domain-specific metadata
- **Search**: retrieve indexed metadata, including the location and access method if known

FEATURES

- **Integrate:** work with existing research tools or workflows to obtain additional metadata
- **Connect:** allow researchers and data managers to see all datasets within one instance of Radiam
- **Package**: assemble metadata for a dataset to assist with publication to a repository

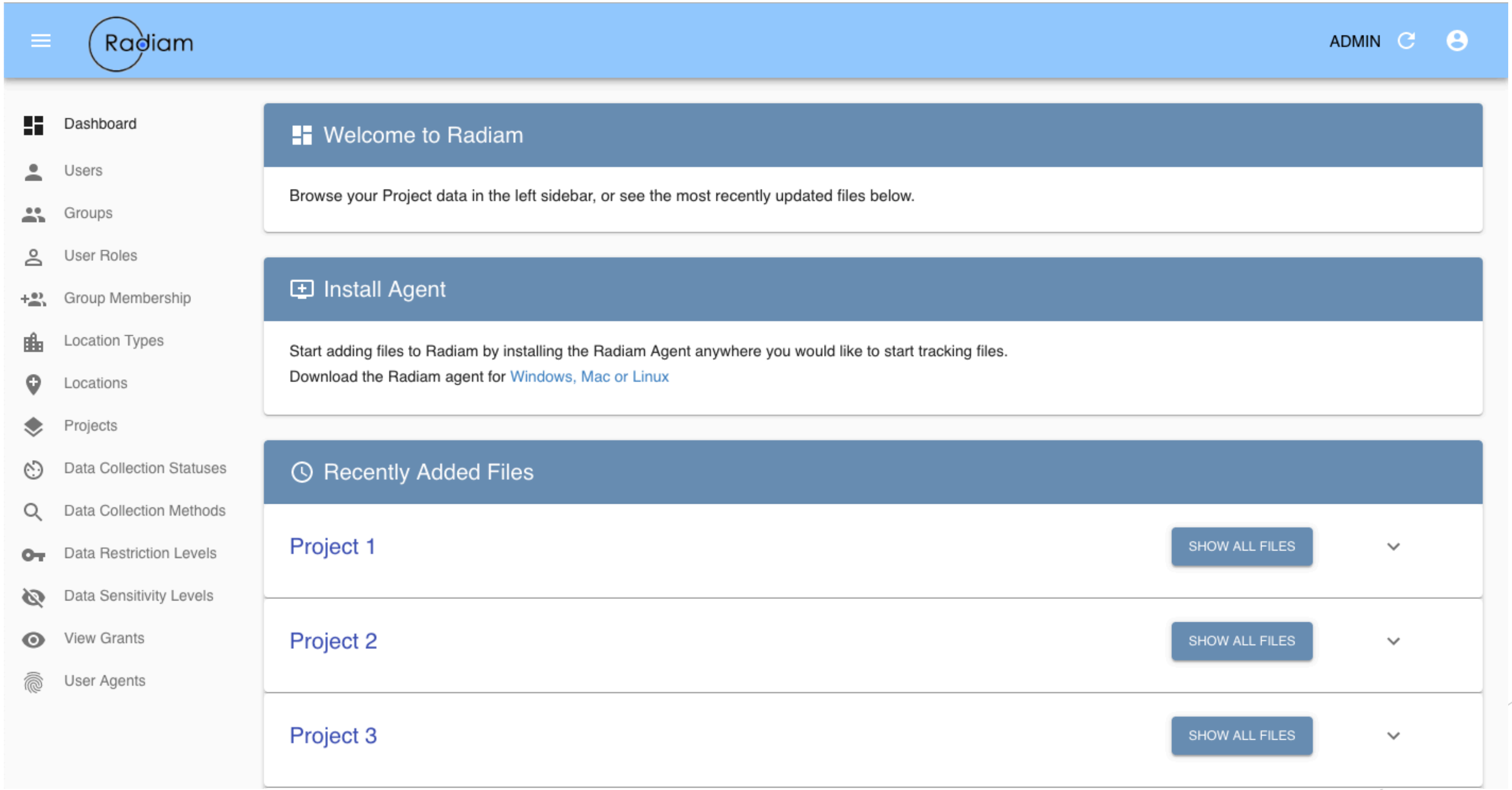
Radiam System Architecture



WHAT’S WORKING NOW

- **API**
  - All REST endpoints to support web and agent functions
  - API specification and documentation
- **Web Interface**
  - Create account, log in, reset password
  - Mange users, groups and projects
  - View the indexed metadata
- **Agent**
  - Built and running on Windows, Mac, Linux
  - Index data for multiple projects with one agent
- **Portal Plugins**
  - HubZero: authentication, search, view the indexed metadata

WEB INTERFACE



Welcome to Radiam

Browse your Project data in the left sidebar, or see the most recently updated files below.

Install Agent

Start adding files to Radiam by installing the Radiam Agent anywhere you would like to start tracking files.
Download the Radiam agent for Windows, Mac or Linux

Recently Added Files

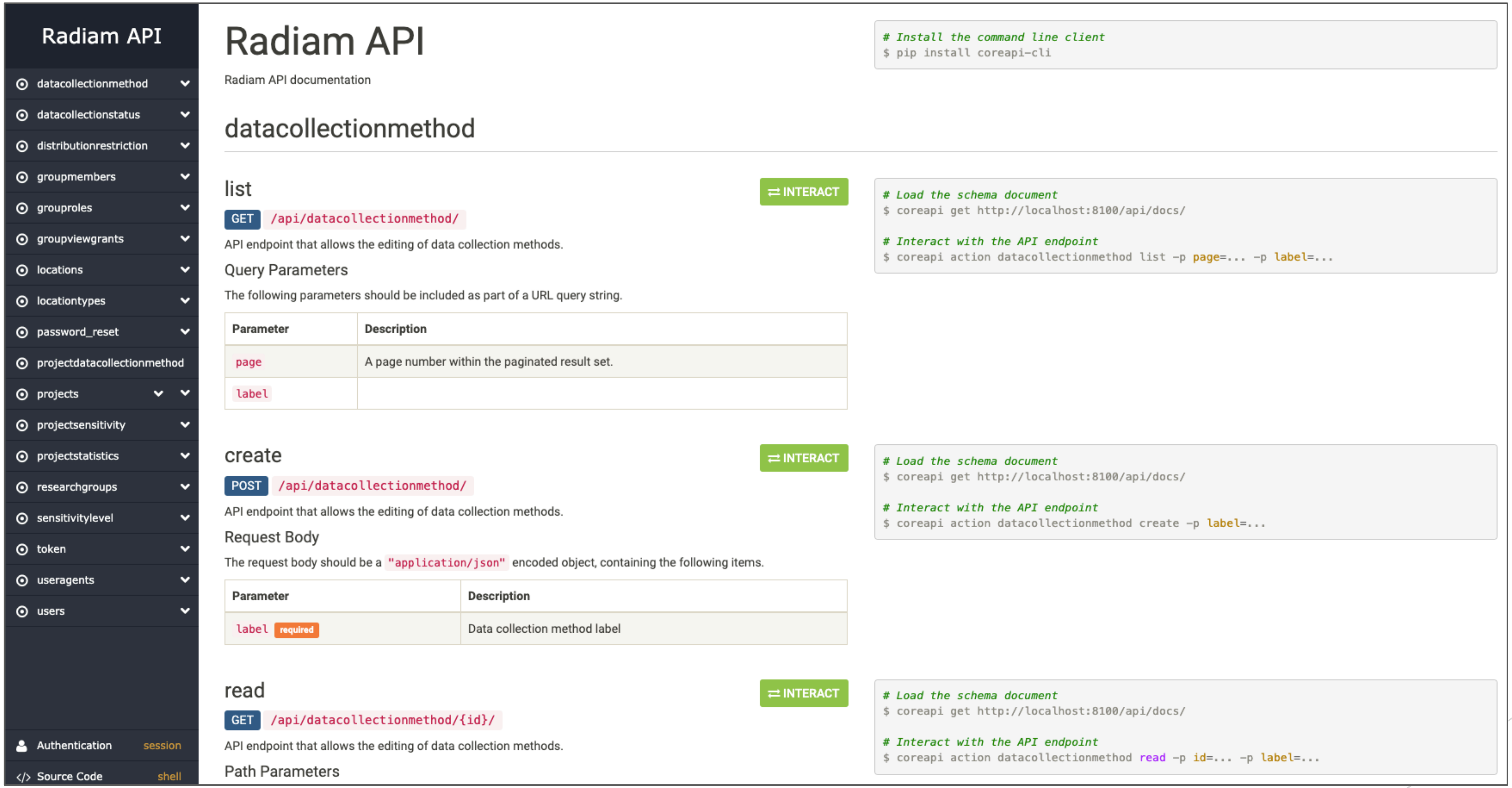
Project 1 SHOW ALL FILES

Project 2 SHOW ALL FILES

Project 3 SHOW ALL FILES

Early draft, subject to change

API



Radiam API

Radiam API documentation

```
# Install the command line client
$ pip install coreapi-cli
```

datacollectionmethod

list

GET `/api/datacollectionmethod/`

API endpoint that allows the editing of data collection methods.

Query Parameters

The following parameters should be included as part of a URL query string.

Parameter	Description
`page`	A page number within the paginated result set.
`label`	

```
# Load the schema document
$ coreapi get http://localhost:8100/api/docs/

# Interact with the API endpoint
$ coreapi action datacollectionmethod list -p page=... -p label=...
```

create

POST `/api/datacollectionmethod/`

API endpoint that allows the editing of data collection methods.

Request Body

The request body should be a `"application/json"` encoded object, containing the following items.

Parameter	Description
`label` required	Data collection method label

```
# Load the schema document
$ coreapi get http://localhost:8100/api/docs/

# Interact with the API endpoint
$ coreapi action datacollectionmethod create -p label=...
```

read

GET `/api/datacollectionmethod/{id}/`

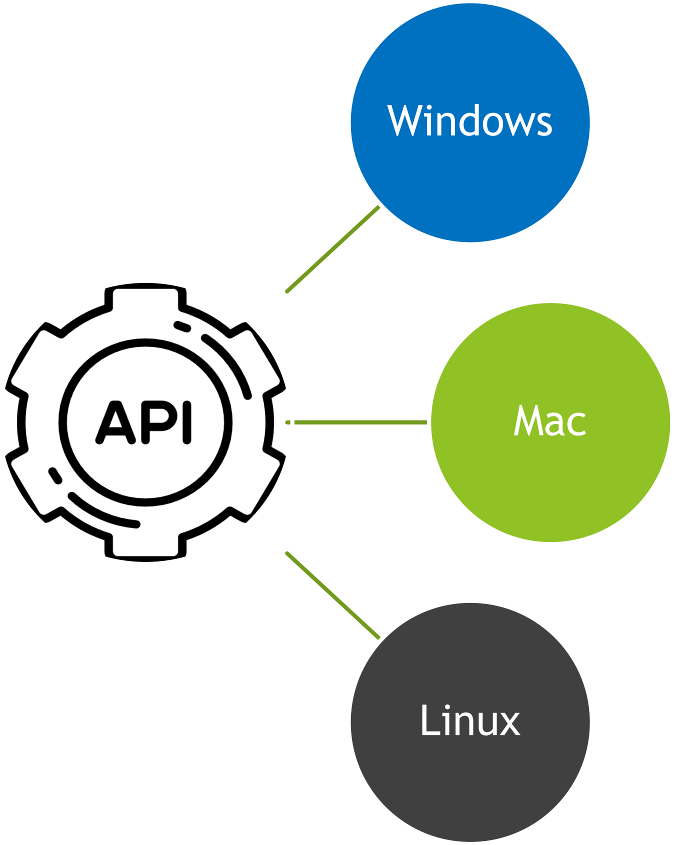
API endpoint that allows the editing of data collection methods.

Path Parameters

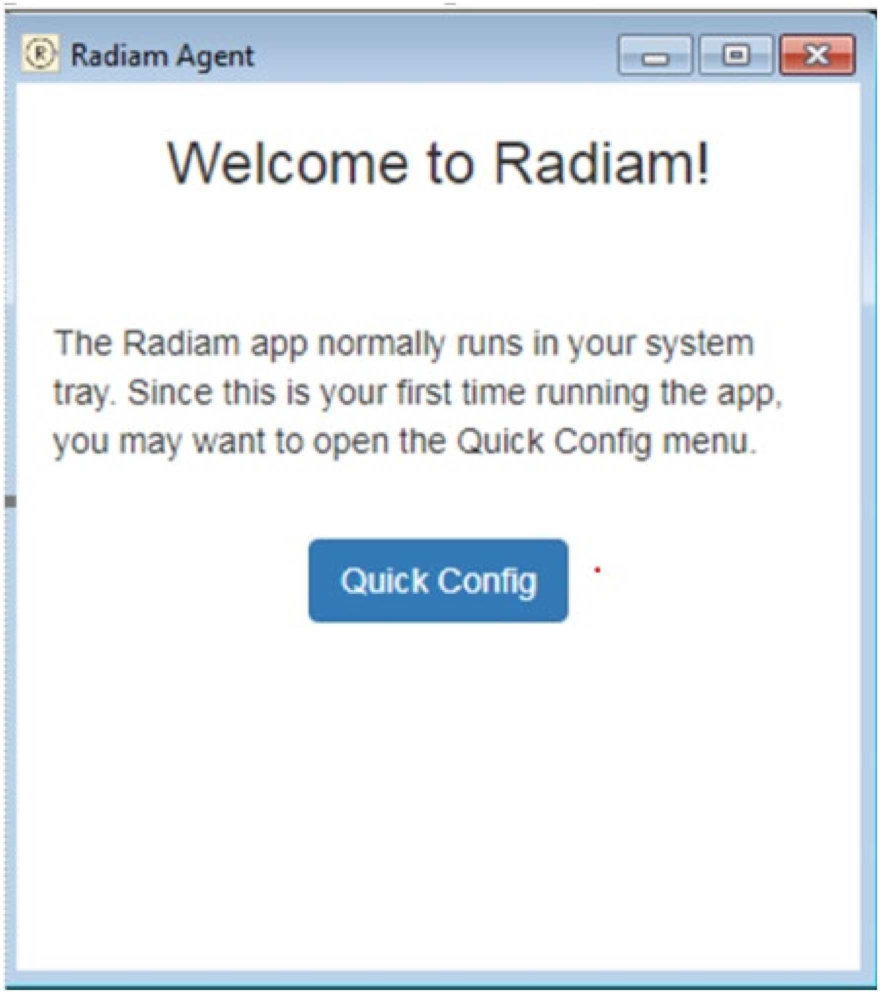
```
# Load the schema document
$ coreapi get http://localhost:8100/api/docs/

# Interact with the API endpoint
$ coreapi action datacollectionmethod read -p id=... -p label=...
```

AGENT



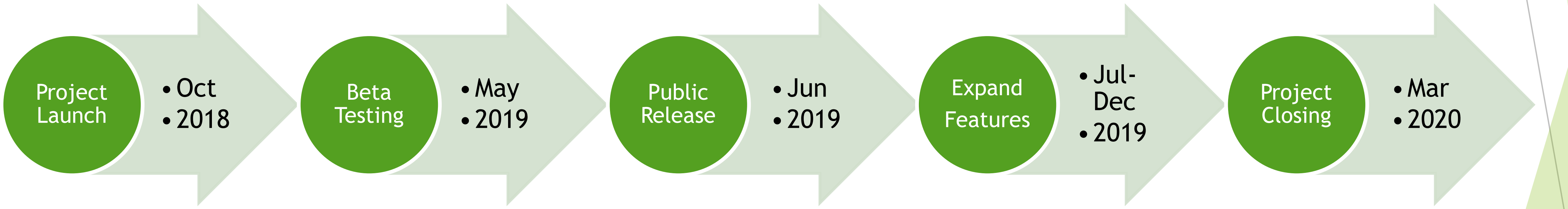
Built and running on 3 platforms



Welcome to Radiam!

The Radiam app normally runs in your system tray. Since this is your first time running the app, you may want to open the Quick Config menu.

TIMELINE



Project Launch
- Oct
- 2018

Beta Testing
- May
- 2019

Public Release
- Jun
- 2019

Expand Features
- Jul-Dec
- 2019

Project Closing
- Mar
- 2020

FUTURE

- Source code to all components of Radiam are being published under the MIT open source license
- The open architecture of Radiam allows its components to be upgraded or rewritten to keep up with integration points such as data portals
- API specification and developer documentation together will allow research groups to write custom applications that work with Radiam
- Published on the CANARIE Research Software Portal: https://science.canarie.ca

PROJECT MEMBERS

- PI: Kevin Schneider, University of Saskatchewan
  - Kevin.Schneider@usask.ca
- Co-PI: Dugan O’Neil, Simon Fraser University
  - doneil@sfu.ca
- Project Lead: Jason Hlady, University of Saskatchewan
  - Jason.hlady@usask.ca
- Project Team:
  - CARL/Portage: Lee Wilson
  - SFU: Alex Garnett, Yang Zhou, Jonathan Loewen
  - USask: Joel Farthing, Todd Trann, Mike Winter, Adam McKenzie, Rama Periasamy, Sergiy Stepanenko