



U N I V E R S I T Y O F

---

L I V E R P O O L

A member of the Russell Group

Python Support for iRODS  
(PyRODS + EmbedPython)

Content Search in iRODS  
(Cheshire3)

John Harrison



UNIVERSITY OF  
LIVERPOOL

# Overview

- Python Support for iRODS
  - PyRODS
  - EmbedPython
- Content Search for iRODS
  - What is Cheshire3?
  - Indexing and Searching iRODS data
  - Upcoming improvements

# Python Support for iRODS

PyRODS + EmbedPython

- PyRODS
  - Python API for iRODS
- EmbedPython
  - iRODS module to execute microservices written in Python

# Python Support for iRODS

PyRODS + EmbedPython

## Jérôme's Slides

# Cheshire3

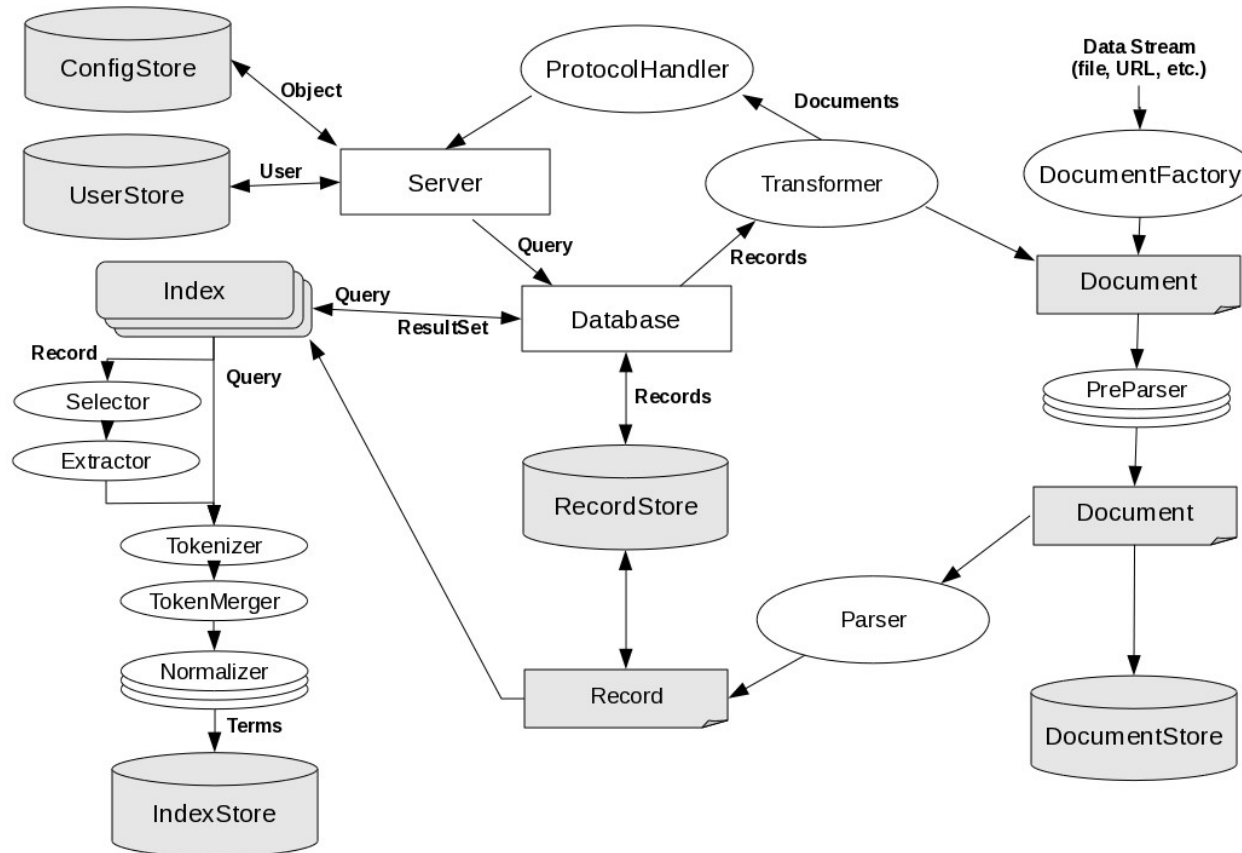
What is it?

Fast XML search engine and framework for information management

- Written in Python
- Flexible XML configuration for indexes to: Browse, Search, Sort search results
- Standard query language: CQL
- Expose data via: SRU, OAI-PMH
- Federated search across SRU targets
- Installed with `easy_install`, `pip` or from source
- The reason for PyRODS + EmbedPython; we wanted to add discovery tools for data stored in iRODS

# Cheshire3

What is it?



# Cheshire3

## Command-line UI

Out-of-the-box database creation, indexing of text-based files (XML, HTML etc.) and search using command-line UI

- **cheshire3-init**
  - ]\$ `cheshire3-init`
- **cheshire3-load**
  - ]\$ `cheshire3-load path/to/file.xml`
  - ]\$ `cheshire3-load path/to/directory`
  - ]\$ `cheshire3-load http://www.example.com/index.html`
- **cheshire3-search**
  - `cheshire3-search "irods"`
  - `cheshire3-search 'dc.title =/word irods and dc.creator exact "john harrison" sortBy dc.creationDate'`



# Indexing and Searching data in iRODS

Today

**IF** you have iRODS, PyRODS + Cheshire3 installed (in the same Python install)...

- Load data from iRODS using the command-line UI
  - ]\$ iinit
  - ... enter credentials
  - ]\$ cheshire3-load -f ifile /zone/home/user/file.xml
  - ]\$ cheshire3-load -f idir /zone/home/user/collection
- Configure Cheshire3 to push Records and Indexes back into iRODS

# Indexing and Searching data in iRODS

Next version (1.1) after release of PyRODS 3.2

- No need for local iRODS installation, automatic resolution of dependencies (incl PyRODS)
  - ]\$ pip install cheshire3[irods]>=1.1
- Command-line UI for creating databases, loading data, and searching in iRODS
  - ]\$ icheshire3-init
  - ]\$ icheshire3-load path/to/collection
  - ]\$ icheshire3-search "irods"
- Other enhancements not-specific to iRODS
  - Improved out-of-the-box support for indexing: PDF, Microsoft Office >= 2007, Open Document Format
  - Natural Language Processing (e.g. identification of Named Entities)

Due for release May 2013

# Running on the iRODS Server

`msiGetStdInExecCmdOut`

Use `msiGetStderrInExecCmdOut`

```
acPostProcForPut |  
msiGetStderrInExecCmdOut(icheshire3-load  
$objPath, *returnVal) |  
nop###nop###nop
```

For more complex workflows, you can write a Python script using the PyRODS and Cheshire3 APIs and call it using EmbedPython

# Questions?

- PyRODS, EmbedPython
  - <http://code.google.com/p/irodspython/>
  - [jerome.fuselier@free.fr](mailto:jerome.fuselier@free.fr)
- Cheshire3
  - <http://cheshire3.org/>
  - <https://github.com/cheshire3/cheshire3>
  - [john.harrison@liv.ac.uk](mailto:john.harrison@liv.ac.uk)

Thank You!