

# iRODS Primer 2

Integrated Rule-Oriented Data System

# Synthesis Lectures on Information Concepts, Retrieval, and Services

## Editor

**Gary Marchionini**, *University of North Carolina, Chapel Hill*

**Synthesis Lectures on Information Concepts, Retrieval, and Services** publishes short books on topics pertaining to information science and applications of technology to information discovery, production, distribution, and management. Potential topics include: data models, indexing theory and algorithms, classification, information architecture, information economics, privacy and identity, scholarly communication, bibliometrics and webometrics, personal information management, human information behavior, digital libraries, archives and preservation, cultural informatics, information retrieval evaluation, data fusion, relevance feedback, recommendation systems, question answering, natural language processing for retrieval, text summarization, multimedia retrieval, multilingual retrieval, and exploratory search.

## **iRODS Primer 2: Integrated Rule-Oriented Data System**

Hao Xu, Terrell Russell, Jason Coposky, Arcot Rajasekar, Reagan Moore, Antoine de Torcy, Michael Wan, Wayne Shroeder, and Sheau-Yen Chen

2017

## **Fuzzy Information Retrieval**

Donald H. Kraft and Erin Colvin

2017

## **Quantifying Research Integrity**

Michael Seadle

2016

## **Incidental Exposure to Online News**

Borchuluun Yadamsuren and Sanda Erdelez

2016

## **Web Indicators for Research Evaluation: A Practical Guide**

Michael Thelwall

2016

### Trustworthy Policies for Distributed Repositories

Reagan W. Moore, Hao Xu, Mike Conway, Arcot Rajasekar, Jon Crabtree, and Helen Tibbo  
2016

### The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really?

Tefko Saracevic  
2016

### Dynamic Information Retrieval Modeling

Grace Hui Yang, Marc Sloan, and Jun Wang  
2016

### Learning from Multiple Social Networks

Liqiang Nie, Xuemeng Song, and Tat-Seng Chua  
2016

### Scholarly Collaboration on the Academic Social Web

Daqing He and Wei Jeng  
2016

### Scalability Challenges in Web Search Engines

B. Barla Cambazoglu and Ricardo Baeza-Yates  
2015

### Social Informatics Evolving

Prina Fichman, Madelyn R. Sanfilippo, and Howard Rosenbaum  
2015

### On the Efficient Determination of Most Near Neighbors: Horseshoes, Hand Grenades, Web Search and Other Situations When Close Is Close Enough, Second Edition

Mark S. Manasse  
2015

### Building a Better World with Our Information: The Future of Personal Information Management, Part 3

William Jones  
2015

### Click Models for Web Search

Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke  
2015

### Information Communication

Feicheng Ma  
2015

## Social Media and Library Services

Lorri Mon

2015

## Analysis and Visualization of Citation Networks

Dangzhi Zhao and Andreas Strotmann

2015

## The Taxobook: Applications, Implementation, and Integration in Search: Part 3 of a 3-Part Series

Marjorie M.K. Hlava

2014

## The Taxobook: Principles and Practices of Building Taxonomies, Part 2 of a 3-Part Series

Marjorie M.K. Hlava

2014

## Measuring User Engagement

Mounia Lalmas, Heather O'Brien, and Elad Yom-Tov

2014

## The Taxobook: History, Theories, and Concepts of Knowledge Organization, Part 1 of a 3-Part Series

Marjorie M.K. Hlava

2014

## Children's Internet Search: Using Roles to Understand Children's Search Behavior

Elizabeth Foss and Allison Druin

2014

## Digital Library Technologies: Complex Objects, Annotation, Ontologies, Classification, Extraction, and Security

Edward A. Fox and Ricardo da Silva Torres

2014

## Digital Libraries Applications: CBIR, Education, Social Networks, eScience/Simulation, and GIS

Edward A. Fox and Jonathan P. Leidig

2014

## Information and Human Values

Kenneth R. Fleischmann

2013

## Multiculturalism and Information and Communication Technology

Pnina Fichman and Madelyn R. Sanfilippo

2013

## Transforming Technologies to Manage Our Information: The Future of Personal Information Management, Part II

William Jones

2013

## Designing for Digital Reading

Jennifer Pearson, George Buchanan, and Harold Thimbleby

2013

## Information Retrieval Models: Foundations and Relationships

Thomas Røelleke

2013

## Key Issues Regarding Digital Libraries: Evaluation and Integration

Rao Shen, Marcos Andre Goncalves, and Edward A. Fox

2013

## Visual Information Retrieval using Java and LIRE

Mathias Lux and Oge Marques

2013

## On the Efficient Determination of Most Near Neighbors: Horseshoes, Hand Grenades, Web Search and Other Situations When Close is Close Enough

Mark S. Manasse

2012

## The Answer Machine

Susan E. Feldman

2012

## Theoretical Foundations for Digital Libraries: The 5S (Societies, Scenarios, Spaces, Structures, Streams) Approach

Edward A. Fox, Marcos André Gonçalves, and Rao Shen

2012

## The Future of Personal Information Management, Part I: Our Information, Always and Forever

William Jones

2012

## Search User Interface Design

Max L. Wilson

2011

## Information Retrieval Evaluation

Donna Harman

2011

**Knowledge Management (KM) Processes in Organizations: Theoretical Foundations and Practice**

Claire R. McInerney and Michael E. D. Koenig  
2011

**Search-based Applications: At the Confluence of Search and Database Technologies**

Gregory Grefenstette and Laura Wilber  
2010

**Information Concepts: From Books to Cyberspace Identities**

Gary Marchionini  
2010

**Estimating the Query Difficulty for Information Retrieval**

David Carmel and Elad Yom-Tov  
2010

**iRODS Primer: Integrated Rule-Oriented Data System**

Arcot Rajasekar, Reagan Moore, Chien-Yi Hou, Christopher A. Lee, Richard Marciano, Antoine de Torcy, Michael Wan, Wayne Schroeder, Sheau-Yen Chen, Lucas Gilbert, Paul Tooby, and Bing Zhu  
2010

**Collaborative Web Search: Who, What, Where, When, and Why**

Meredith Ringel Morris and Jaime Teevan  
2009

**Multimedia Information Retrieval**

Stefan Rüger  
2009

**Online Multiplayer Games**

William Sims Bainbridge  
2009

**Information Architecture: The Design and Integration of Information Spaces**

Wei Ding and Xia Lin  
2009

**Reading and Writing the Electronic Book**

Catherine C. Marshall  
2009

**Hypermedia Genes: An Evolutionary Perspective on Concepts, Models, and Architectures**

Nuno M. Guimarães and Luís M. Carrico  
2009

## Understanding User-web Interactions via Web Analytics

Bernard J. (Jim) Jansen

2009

## XML Retrieval

Mounia Lalmas

2009

## Faceted Search

Daniel Tunkelang

2009

## Introduction to Webometrics: Quantitative Web Research for the Social Sciences

Michael Thelwall

2009

## Exploratory Search: Beyond the Query-response Paradigm

Ryen W. White and Resa A. Roth

2009

## New Concepts in Digital Reference

R. David Lankes

2009

## Automated Metadata in Multimedia Information Systems: Creation, Refinement, Use in Surrogates, and Evaluation

Michael G. Christel

2009

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2017

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

iRODS Primer 2: Integrated Rule-Oriented Data System

Hao Xu, Terrell Russell, Jason Coposky, Arcot Rajasekar, Reagan Moore, Antoine de Torcy,  
Michael Wan, Wayne Shroeder, and Sheau-Yen Chen

ISBN: 978-3-031-01181-8      paperback

ISBN: 978-3-031-02309-5      ebook

DOI 10.1007/978-3-031-02309-5

A Publication in the Springer series

*SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL, AND SERVICES*

Lecture #57

Series Editor: Gary Marchionini, *University of North Carolina, Chapel Hill*

Series ISSN

Print 1947-945X    Electronic 1947-9468



# iRODS Primer 2

## Integrated Rule-Oriented Data System

Hao Xu, Terrell Russell, Jason Coposky, Arcot Rajasekar, Reagan Moore, and  
Antoine de Torcy  
University of North Carolina at Chapel Hill

Michael Wan, Wayne Shroeder, and Sheau-Yen Chen  
University of California, San Diego

*SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL,  
AND SERVICES #57*

## ABSTRACT

Policy-based data management enables the creation of community-specific collections. Every collection is created for a purpose. The purpose defines the set of properties that will be associated with the collection. The properties are enforced by management policies that control the execution of procedures that are applied whenever data are ingested or accessed. The procedures generate state information that defines the outcome of enforcing the management policy. The state information can be queried to validate assessment criteria and verify that the required collection properties have been conserved. The integrated Rule-Oriented Data System implements the data management framework required to support policy-based data management. Policies are turned into computer actionable Rules. Procedures are composed from a microservice-oriented architecture. The result is a highly extensible and tunable system that can enforce management policies, automate administrative tasks, and periodically validate assessment criteria. iRODS 4.0+ represents a major effort to analyze, harden, and package iRODS for sustainability, modularization, security, and testability. This has led to a fairly significant refactorization of much of the underlying codebase. iRODS has been modularized whereby existing iRODS 3.x functionality has been replaced and provided by small, interoperable plugins. The core is designed to be as immutable as possible and serve as a bus for handling the internal logic of the business of iRODS. Seven major interfaces have been exposed by the core and allow extensibility and separation of functionality into plugins.

## KEYWORDS

data life cycle, data grid, digital library, preservation environment, policy-based data management, rule engine, iRODS, metadata catalog, assessment criteria, policies, microservices

# Contents

	<b>Acknowledgments</b> .....	<b>xv</b>
<b>1</b>	<b>Introduction</b> .....	<b>1</b>
<b>2</b>	<b>Integrated Rule-Oriented Data System</b> .....	<b>3</b>
2.1	Data Grid Overview .....	4
<b>3</b>	<b>iRODS Architecture</b> .....	<b>13</b>
3.1	Virtualization in iRODS .....	17
3.2	iRODS Components .....	18
3.3	User Environment Variables .....	19
3.4	Configuration Files .....	19
3.4.1	~/odbc.ini .....	19
3.4.2	~/irods/.irodsA .....	19
3.4.3	/etc/irods/server_config.json .....	20
3.4.4	~/irods/irods_environment.json .....	23
3.4.5	Checksum Configuration .....	25
3.4.6	Special Characters .....	26
3.5	Plugin Interfaces .....	26
3.5.1	Pluggable Microservices .....	27
3.5.2	Composable Resources .....	27
3.5.3	Pluggable Authentication .....	28
3.5.4	Pluggable Network .....	28
3.5.5	Pluggable Database .....	29
3.5.6	Pluggable RPC API .....	29
3.5.7	Pluggable Rule Engine .....	30
3.6	Example Plugins .....	30
3.6.1	Composable Resources .....	30
3.6.2	Pluggable Authentication .....	40
<b>4</b>	<b>Rule-Oriented Programming</b> .....	<b>51</b>
4.1	Session State Variables .....	55
4.2	Persistent State Information Variables .....	55

<b>5</b>	<b>The iRODS Rule System</b>	<b>57</b>
5.1	The iRODS Rule Architecture	59
5.2	Default iRODS Rules	60
5.3	Session Variables Available for Default iRODS Rules	61
5.4	Dynamic Policy Enforcement Points	63
5.4.1	Flow Control	64
5.4.2	Parameter Serialization	64
5.4.3	Dynamic PEP Signatures	65
5.5	Pluggable Rule Engine Architecture	67
5.5.1	Rule Engine Plugin Framework	67
5.5.2	Configuration (server_config.json)	67
5.6	The iRODS Rule Language	69
5.6.1	Comments	69
5.6.2	Directives	70
5.6.3	Boolean and Numeric	70
5.6.4	Strings	72
5.6.5	Dot Expression	76
5.6.6	Constant	76
5.6.7	Function	77
5.6.8	Rule	77
5.6.9	Data Types and Pattern Matching	79
5.6.10	Control Structures	83
5.6.11	Recovery Chain For Control Structures	86
5.6.12	Types	88
5.6.13	Language Integrated General Query	93
5.6.14	Path Literals	94
5.7	Delay Execution	95
5.7.1	Syntax	95
5.7.2	Examples	96
5.8	Remote Execution	97
5.8.1	Syntax	97
5.8.2	Examples	97
<b>6</b>	<b>iRODS Microservices</b>	<b>99</b>
6.1	Microservice Input/Output Arguments	102
6.2	About Microservices and Microservice Plugins	102

6.3	Examples .....	103
6.3.1	The Plugin Factory .....	104
6.3.2	The Microservice Definition .....	105
6.3.3	Building and Installing the Example Code .....	107
6.3.4	Testing the Microservice .....	107
<b>A</b>	<b>Exercises .....</b>	<b>109</b>
A.1	Short Questions .....	109
A.2	Essay Questions .....	111
	<b>Authors' Biographies .....</b>	<b>113</b>

# Acknowledgments

This research was supported by:

- NSF ITR 0427196, Constraint-Based Knowledge Systems for Grids, Digital Libraries, and Persistent Archives (2004–2007)
- NARA supplement to NSF SCI 0438741, Cyberinfrastructure; From Vision to Reality—Developing Scalable Data Management Infrastructure in a Data Grid-Enabled Digital Library System (2005–2006)
- NARA supplement to NSF SCI 0438741, Cyberinfrastructure; From Vision to Reality—Research Prototype Persistent Archive Extension (2006–2007)
- NSF SDCI 0721400, SDCI Data Improvement: Data Grids for Community Driven Applications (2007–2010)
- NSF/NARA OCI-0848296, NARA Transcontinental Persistent Archive Prototype (2008–2012)

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Archives and Records Administration, the National Science Foundation, or the U.S. Government.

Hao Xu, Terrell Russell, Jason Coposky, Arcot Rajasekar, Reagan Moore, Antoine de Torcy, Michael Wan, Wayne Shroeder, and Sheau-Yen Chen  
March 2017