



Introduction to the USRN Discovery Toolkit

Paul Walk

Director & Founder, Antleaf
paul@paulwalk.net

•ANTLEAF•
www.antleaf.com

Contents

- Part 1: Introduction to the Toolkit
- Part 2: Demonstration
- Part 3: Feedback - and invitation to participate

Part 1: Introduction

Overview

- The Toolkit is a **wiki** containing resources to support the USRN Discovery project
- Helping USRN repositories to:
 - increase interoperability across the national and global repository network
 - establish suitability of a digital publication repository to be an "agency-designated" repository for federally funded research
- where the concept of an "agency-designated" repository is to be understood in the context of the OSTP "Nelson Memo"

Toolkit

paulwalk edited this page now - 18 revisions

This "toolkit" is designed to help repositories to adopt some of the "desirable characteristics" already identified by USRN (see [source documents](#)).

Desirable characteristics

The "desirable characteristics" for repositories taken from the USRN document [Desirable Characteristics of Digital Publication Repositories](#) are listed below. Each links to an information page which describes the characteristic and links to corresponding characteristics from the [COAR Community Framework for Good Practices in Repositories](#).

A subset of the desirable characteristics are the focus of this toolkit - these are listed first. The USRN Discovery project team have developed some adoption strategies for each of these.

Characteristics in focus for USRN Discovery

(these have adoption strategies developed by the USRN Discovery project team)

- [Free and Easy Discoverability and Access](#)
- [Clear Use Guidance](#)
- [Unique Persistent Identifiers](#)
- [Metadata](#)
- [Broad and Measured Reuse](#)
- [Common Format](#)

Other desirable characteristics

- [Preservation](#)
- [Curation and quality assurance](#)
- [Long-term Organizational Sustainability](#)
- [Provenance](#)
- [Authentication](#)
- [Long-term Technical sustainability](#)
- [Security and Integrity](#)

General Resources

The following resources are more generally useful in terms of adopting good practice in the management of repositories.

- [Ten Recommended Practices for Managing Preprints in Generalist and Institutional Repositories](#)

Providing feedback

You can provide feedback on, or report an issue with, this Toolkit, by [creating a GitHub issue](#).

Pages 21

Home

Toolkit

- [Toolkit start page](#)
- [Free and Easy Discoverability and Access](#)
- [Clear Use Guidance](#)
- [Unique Persistent Identifiers](#)
- [Metadata](#)
- [Broad and Measured Reuse](#)
- [Common Format](#)
- [Resources supporting the Toolkit](#)
 - [Riox metadata format](#)
 - [Controlled Vocabularies](#)
- [Providing feedback on the Toolkit](#)
 - [Feedback](#)
- [USRN Discovery Project](#)
 - [USRN Discovery: Participating repositories](#)

Clone this wiki locally

<https://github.com/antleaf/USRN-Toolkit>

<https://bit.ly/USRN-toolkit>

Structure of the Toolkit

- Framed by *Desirable Characteristics of Digital Publication Repositories*, [USRN, 2023]
- The Toolkit deals specifically with a subset (6/13) of these "desirable characteristics":
 - Free and Easy Discoverability and Access
 - Clear Use Guidance
 - Unique Persistent Identifiers
 - Metadata
 - Broad and Measured Reuse
 - Common Format

U.S. Repository Network

Desirable Characteristics of Digital Publication Repositories

Version 1.0

Approved March 31, 2023

Introduction

A set of desirable characteristics alongside good practices is necessary to

- increase interoperability across the national and global repository network and
- establish suitability of a digital publication repository to be an "agency-designated" repository for federally funded research.¹

These desirable characteristics are intended to align with / work in parallel with the already issued *Desirable Characteristics of Data Repositories for Federally Funded Research*² and are modeled on three existing frameworks.

1. [NSTC Desirable Characteristics of Data Repositories for Federally Funded Research](#)
2. [COAR Community Framework for Good Practices in Repositories, Version 2, Essential Characteristics](#)
3. [Plan S Requirements for Open Access Repositories, Mandatory Criteria](#)

In defining the desirable characteristics outlined below, the U.S. Repository Network (USRN) Desirable Characteristics Working Group strove to balance feasibility and necessity. Not all U.S. repositories currently possess all the resulting desirable characteristics, and, in some cases,

Structure of the Toolkit (2)

- Each of the "desirable characteristics" has a page with:
 - Description
 - USRN Adoption Strategies
 - Descriptions of approaches to adopting this "desirable characteristic"
 - Check your repository in the CORE Dashboard
 - Information about and direct links to the relevant CORE Dashboard tools
 - Related characteristics from the COAR Community Framework
 - Relationship to a complementary framework from COAR
 - Relevant Resources in this Toolkit
 - Links to related pages in the Toolkit
 - Relevant External Resources
 - Links to relevant resources outwith the Toolkit

External Resources

- Some significant external resources referenced in the Toolkit:
 - COAR Framework
 - COAR Vocabularies
 - RIOXX
 - (Matteo will say more)

COAR Confederation of Open Access Repositories

News & Updates What we do Resources Members About COAR Search

COAR Community Framework for Good Practices in Repositories

Purpose

The aim of the framework is to assist repositories to evaluate and improve their current operations based on a set of applicable and achievable good practices.

There are a number of existing frameworks and tools available to support repositories in improving various facets of their operations. However, many of these have been developed for a specific type of repository (e.g., institutional or disciplinary). This framework aims to have a more comprehensive and generic approach, applicable to all types of repositories around the world.

The COAR Community Framework provides a set of guidelines and best practices for different types of repositories (publication, institutional, disciplinary, etc.)

Process

The framework was developed and is maintained by a diverse group of experts from around the world, representing different regions and variety of institutions. The framework includes a process for developing, maintaining, and updating the framework, which were supplemented and reviewed by a panel of experts.

The 2nd version of the COAR Community Framework was released in 2022, with plans for publishing an updated version in late 2024.

The COAR Community Framework

The framework contains 54 criteria organized into six main sections: authenticity, quality assurance, preservation, access, user experience, and characteristics and desired characteristics, based on the COAR Principles of Good Practice.

[Download "The COAR Community Framework"](#)

COAR Controlled Vocabularies for Repositories

Home Access Rights Resource Types Version Types About Give Feedback

Resource Types 3.1

This is the latest version of this vocabulary

Concepts	Description
<ul style="list-style-type: none">◦ cartographic material<ul style="list-style-type: none">▪ map▪ dataset◦ aggregated data◦ clinical trial data◦ compiled data◦ encoded data◦ experimental data◦ genomic data◦ geospatial data◦ laboratory notebook◦ measurement and test data◦ observational data◦ recorded data◦ simulation data◦ survey data	The Resource Type vocabulary defines concepts to identify the genre of a resource. Such resources, like publications, research data, audio and video objects, are typically deposited in institutional and thematic repositories or published in ejournals. This vocabulary supports a hierarchical model that relates narrower and broader concepts. Multilingual labels regard regional distinctions in language and term. Concepts of this vocabulary are mapped with terms and concepts of similar vocabularies and dictionaries.
<h3>Creators</h3> <ul style="list-style-type: none">• Cristina Azorin (Catalan translation)• Paola Azrilevich (Spanish translation)• Isabel Bernal (Spanish translation)• Gültækir Gürdal (Turkish translation)• Sawsan Habre (Arabic translation)• Juha Hakala (Finnish translation)• Yutaka Hayashi (Japanese translation)• İlkyay Holt (Turkish translation)• Nie Hua (Chinese translation)• Hilary Jones (English translation)	<h3>Released</h3> 2022-09-29
<h3>Namespace URI</h3> http://purl.org/coar/resource_type/	<h3>Creators</h3> <ul style="list-style-type: none">• Cristina Azorin (Catalan translation)• Paola Azrilevich (Spanish translation)• Isabel Bernal (Spanish translation)• Gültækir Gürdal (Turkish translation)• Sawsan Habre (Arabic translation)• Juha Hakala (Finnish translation)• Yutaka Hayashi (Japanese translation)• İlkyay Holt (Turkish translation)• Nie Hua (Chinese translation)• Hilary Jones (English translation)
<h3>Vocabulary PID</h3> https://vocabularies.coar-repositories.org/resource_types/	<h3>Versions</h3> <ul style="list-style-type: none">• 3.1 (latest version)

Rioxx

The Research Outputs Metadata Schema

Rioxx 3.0

This is the current version of the application profile
[see other versions](#)

Terminology

the primary resource refers to the thing being described by the Rioxx record.

the resource refers to resources that are not the *primary resource*. A resource can encompass resource identifiers (i.e. URIs), such as persistent identifiers.

landing page refers to an entry page on a repository or publication platform that describes the *primary resource* and provides links to related resources.

custodianship refers to the concept that the location of a *primary resource* and a resource is significant to understanding the nature of the thing being described by the Rioxx record and is relevant to machine processing of Rioxx metadata. Resources under *direct custodianship* are those which are under direct management of a local repository or publication platform, and under which that local system controls and maintains file content and/or other resources. Resources outside this *direct custodianship* (e.g. resources at or hosted by third party services) demonstrate *external custodianship*.

The terms **MUST**, **MUST NOT**, **REQUIRED**, **SHALL**, **SHALL NOT**, **SHOULD**, **SHOULD NOT**, **RECOMMENDED**, **MAY**, and **OPTIONAL** used in the table below should be interpreted as described in [RFC 2119](#).

Properties in this profile

[dc:coverage](#) | [dc:description](#) | [dc:identifier](#) | [dc:language](#) | [dc:relation](#) | [dc:source](#) | [dc:subject](#) | [dc:title](#) | [dc:type](#) | [dcterms:dateAccepted](#) | [rioxxterms:contributor](#) | [rioxxterms:creator](#) | [rioxxterms:ext_relation](#) | [rioxxterms:grant](#) | [rioxxterms:id](#) | [rioxxterms:name](#) | [rioxxterms:project](#) | [rioxxterms:publication_date](#) | [rioxxterms:publisher](#) | [rioxxterms:record_public_release_date](#)

Example Rioxx Records

Some examples of full Rioxx records (serialised as XML) are provided at the bottom of this page.

Property details and examples

Element	Cardinality	Description
dc:coverage	Zero or more	Coverage (dc:coverage) will typically include a temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity).
dc:description	Zero or more	The purpose of dc:description is to describe the content of the <i>primary resource</i> using free text. It is RECOMMENDED that an English language abstract be used where available. HTML or other markup tags SHOULD NOT be included in this field. This field may be indexed and its contents presented to people conducting searches.
dc:identifier	Exactly one	dc:identifier MUST contain an identifying HTTP(S) URI of the repository 'landing page' describing this resource. The landing page should provide links to related resources. We RECOMMEND using FAIR Signposting (signposting.org/FAIR) to assist machine agents in navigating from the landing page to the relevant resources. The value of dc:identifier does not have to be a persistent identifier (PID), instead, in order to facilitate indexing by machine agents, it is RECOMMENDED that dc:identifier contains a direct HTTP(S) URI of the landing page, i.e. resolving dc:identifier SHOULD be possible without the use of HTTP redirection.

Editorial Control

- The Toolkit has been developed by the USRN Discovery Pilot team
- **However**, we intend it to be owned by the USRN
- We are inviting the participants in the USRN Discovery Pilot to participate in adding to, and maintaining the Toolkit
 - Basically, this involves editing the wiki.



Part 2: Demonstration

Part 3: Feedback & participation

Accessing the USRN Toolkit

- "Read" access
 - The Toolkit is completely open for public read access
- "Write" access
 - To edit or create pages in the Toolkit, you will need a (free) Github account, to which the USRN Discovery Pilot team will then grant write permissions
- Providing feedback:
 - To provide feedback or report an issue, you can raise a GitHub issue (the Toolkit includes directions for this)

<https://bit.ly/usrn-toolkit>