DReSA Automated Ingestion Provider's Guide

Version 1.4.2

Contents

1. About DReSA

DReSA began as a grassroots movement among trainers and training providers in Australia and New Zealand to improve the discoverability of digital skills training. It was an outcome of the ARDC's Australian eResearch and Data Skills Summit 2020 that led to the establishment of the *National Training Registry and Calendar Working Group*. And, after scoping the landscape for appropriate solutions, the *Working Group* found the ELIXIR TeSS platform, which provides the code base from which DReSA was built. From there a truly collaborative, global partnership with ELIXIR began, with an exchange of both the infrastructure that underlies TeSS, and ideas, support and encouragement.

DReSA seeks to:

- Lift Australasian researchers' digital capabilities and skills by improving the findability of training events and resources.
- Build national, collaborative, training partnerships.
- Strategically leverage trainers' and learners' time and resources.
- make it easier for researchers, trainers, and training providers to find digital research skills-focused educational events and resources.

There is more information available on the following subjects by following the *About* link on the navigation bar:

- What is DReSA?
- How to register content (events, materials, and providers) in DReSA.
- Information for developers who want to connect their training systems to DReSA.
- How to contact the DReSA team, and who are the people and organisations that have worked to establish DReSA.

The rest of this manual will describe the features of the DReSA web application, which is accessible at the following address: https://www.dresa.org.au

2. What is Automated Ingestion

DReSA can provide various automated ingestion solutions that pull resources into the application. These solutions are based on the extraction of structured metadata from existing URLs.

If those solutions are not applicable, you can contact the DReSA team to discuss potential, alternative solutions. This choice is particularly useful if you add many events to DReSA.

3. Ingestion sources

DReSA can be configured to ingest resources from a variety of sources. However, each source must allow DReSA to extract and transform data, or set default values, for the following required fields:

- **Event**: Title, URL, Start, End, Organiser, Description, Host Institutions, Contact, Eligibility, Timezone, Online, City (online is false), and Country (online is false).

- **Material**: Title, URL, Description, Keywords, Licence, Status, Contact

A variety of other fields can be ingested, depending on the ingestion source, but these data must be available to be able to create a resource in DReSA.

Matching resources

The Title and URL fields, along with the Start field (for events only), and the source's provider are used to match with an existing resource in DReSA. If a successful match is found the resource will be updated instead of a new resource being inserted each time the source is ingested. In addition, ingested values will update the fields on matched resources except for any fields that have been marked as 'locked'.

Impact of updates

The potential impacts of ingesting matched resources can vary depending on where an update has occurred, as follows:

- **Source record** has changed: the matched resource in DReSA will be overwritten unless the changed fields have been 'locked' by the resource owner/editor.
- **Matched resource** has changed: as above.
- **Matching fields** have been changed on either the source record or the matched resource: a new resource will be created in DReSA with the attributes of the source record.

To avoid unwanted changes, the provider should take care to lock important fields and avoid changing the values of the matching fields in DReSA, i.e. Title, URL, and Start Time.

Configuring a new source

First, a provider needs to be registered in DReSA. Then, a source can be defined for each URL that is to be accessed for ingestion. The following fields define an ingestion source:

- **provider:** *Title of the provider's DReSA record*
- **url**: *URL of the source*
- **resource_type:** `'event' or 'material'`
- **method**: `'csv', 'ical' (events) or 'rest'`
- **token:** *Private token for the Eventbrite API*

When this data has been provided, any user with *Administrator* role can create an ingestion source record for the provider.

4. Method: iCalendar

This method only applies to the **events** resource type. It supports the download and extraction of event data from files that are formatted using the iCalendar standard [1]. DReSA can be configured to find iCalendar files from a sitemap or directly from a base URL. Sitemaps provide a standard schema that lists appropriate pages within a site (such as: https://dresa.org.au/sitemap.xml).

The *source* can refer to either of the following:

- a 'sitemap.xml' file, from which a set of base URLs can be extracted
  for example: https://pawsey.org.au/tribe_events-sitemap.xml

- a base URL, from which an .ics or .ical file can be downloaded
  for example: https://pawsey.org.au/event/pawsey-clinic-ecu-joondalup/?ical=true
  Note: the query parameter '?ical=true' may be omitted from the source and will be appended in the ingestion process.

The following event fields are not included in an icalendar file:

- **contact** – will be set to the Content Provider's *contact*.
- **organizer** – will be set to the Content Provider's *title*.
- **host institutions** – will be set to the Content Provider's *title*.
- **eligibility** – will be set to '*open_to_all*'.

All these values can be edited and locked once the event has been successfully ingested.

An example .ics file can be found via the following link:
https://github.com/nrmay/TeSS/blob/master/test/fixtures/files/icalendar/ask-me-anything-porous-media-visualisation-and-lbpm.ics

5. Method: REST API

The 'rest' method has currently been implemented for ingestion from the following APIs:

- The ElixirTeSS training platform for ingestion of events.
- The Eventbrite event management platform for ingestion of events.
  A provider needs to create an API Key and provide the private token, so that it can be included for authentication of API queries. See the following link for more information about API Keys: https://www.eventbrite.com/platform/api/#/introduction/authentication
- The Zenodo open-access repository for ingestion of materials.

To ingest materials or events from other REST APIs, a query interface and field mapping will need to be provided, so that resources can be found and extracted.

See the following link for more information about REST APIs: What is REST (restfulapi.net)

[1] RFC 5545: Internet Calendaring and Scheduling Core Object Specification (iCalendar).
Link as follows: https://datatracker.ietf.org/doc/html/rfc5545

Mapping of API attributes

Attributes from the various APIs are mapped into DReSA where names are matched. Other attributes are mapped as follows:

API	Source Attribute	DReSA Attribute
ElixirTeSS	*online* = true or *venue* = 'Online'	*online* = true
	'UTC'	*timezone*
Eventbrite	*online_event* = true	*online* = true
	start.timezone	*timezone*
	start.local	*start*
	end.local	*end*
	name.text	*title*
	description.html	*description*
	organizer.name	*organizer*
	venue.address_1 + ',' + *venue.address_2*	*venue*
	category and *subcategory*	*keywords*
	format.short_name	*event_types*
	invite_only 1. true 2. false, null	*eligibility* 1. by_invitation 2. open_to_all
	is_free 1. true 2. false, null	*cost_basis* 1. free 2. charge
	currency	*cost_currency*
Zenodo	'active'	*status*
	metadata.license.id	*licence*
	metadata.creators: name (orcid)	*authors*
	metadata.contributors: name (type)	*contributors*
	links.html	*url*
	metadata	other attributes

The following fields are mapped by value across APIs:

eligibility	1. first_come_first_served 2. registration_of_interest 3. by_invitation	1. open_to_all 2. expression_of_interest 3. by_invitation
event_types	1. conference 2. class 3. networking 4. meetings_and_conferences 5. workshops_and_courses	1. conference 2. workshop 3. meeting 4. meeting 5. workshop

6. Method: Comma Separated Values (CSV) file

A separate file is required for each type of resource that is being ingested in csv format. The information required to set up the ingestion are the URL of each file and the name of the provider to which they should be linked.

Materials File

This file should have a header row and one row for each material, with columns defined as follows:

header	**format**	**req?**	**notes**
Title	String [2]	Y	
URL	URL	Y	
Description	"""Text"""	Y	Can include markdown [3] or html. Note [4]
Keywords	String;String;	Y	One or more semi-colon separated strings
Contact	String	Y	Defaults to Provider's contact if blank
Licence	String	Y	A valid licence key [5]

Status	String	Y	One of the following keys:
			`active`
			`development`
			`archived`

Optional Fields – the following columns can be included, in any order, or omitted completely.			
DOI	String		
Version	String		
Published	Date		String in format: YYYYmmdd
Modified	Date		String in format: YYYYmmdd

Competency	String		One of the following keys:
			`notspecified`
			`awareness`
			`beginner`
			`intermediate`
			`advanced`

Authors	String;String;		One or more semi-colon separated authors
Contributors	String;String;		One or more semi-colon separated contributors
Fields	String;String;		One or more semi-colon separated Fields of Research [6]
Audiences	String;String;		One or more semi-colon separated audience keys [7]
Types	String;String;		One or more semi-colon separated resource type keys [8]
Other Types	String	*	Other types (required if Types includes key 'other')
Objectives	"""Text"""		Can include markdown or html.
Prerequisites	"""Text"""		Can include markdown or html.
Syllabus	"""Text"""		Can include markdown or html.

[2] A string is text up to a maximum of 255 characters.
[3] See following for reference: https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet
[4] Note: double-quote characters within the text should be replaced with two double-quotes ("").
[5] Licence keys can be found in the following file:
https://github.com/dresa-org-au/TeSS/blob/master/config/dictionaries/target_audience.yml
[6] Browse and search for fields at ARDC – Research Vocabularies Australia: ANZSRC 2008: Fields of Research
[7] Target audience keys can be found in the following file:
https://github.com/dresa-org-au/TeSS/blob/master/config/dictionaries/target_audience.yml
[8] Material type keys can be found in the following file:
https://github.com/dresa-org-au/TeSS/blob/master/config/dictionaries/material_type.yml

Events File

This file should have a header row and one row for each event, with columns defined as follows:

header	format	req?	notes
Title	String [9]	Y	
URL	URL	Y	
Description	"""Text"""	Y	Can include markdown [10] or html. Note [11]
Start	DateTime	Y	String in format: YYYYmmddTHHMMSS
End	DateTime	Y	String in format: YYYYmmddTHHMMSS
Timezone	String	Y	A valid TZ database name [12]
Contact	String	Y	Defaults to the Provider's contact if blank
Organizer	String	Y	

Eligibility	String;String;	Y	One or more of the following keys:
			`open_to_all`
			`host_institution`
			`expression_of_interest`
			`by_invitation`

Host Institutions	String;String;	Y	One or more semi-colon separated strings
Online	String	Y	`true` or `false`
Optional Fields – the following columns can be included, in any order, or omitted completely.			
City	String	*	Required if not online
Country	String	*	Required if not online. Alpha 2 code [13]
Venue	"""Text"""		
Postcode	String		
Subtitle	String		
Duration	String		
Recognition	String		
Types	String;String;		One or more of the following keys: `webinar, workshop, conference, meeting, hackathon, dropin`

Cost Basis	String		One of the following keys:
			`free`
			`hosts`
			`charge`

Currency	String		Three letter alphabetic currency code [14]
Cost	Number		Decimal value
Capacity	Number		Integer greater than zero
Fields	String;String;		One or more semi-colon separated strings [15]
Keywords	String;String;		One or more semi-colon separated strings
Audiences	String;String;		One or more semi-colon separated strings
Objectives	"""Text"""		Can include markdown or html.
Prerequisites	"""Text"""		Can include markdown or html.
Requirements	"""Text"""		Can include markdown or html.

[9] A string is text up to a maximum of 255 characters.
[10] See following for reference: https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet
[11] Note: double-quote characters within the text should be replaced with two double-quotes ("").
[12] The full list of TZ database names: https://en.wikipedia.org/wiki/List_of_tz_database_time_zones
[13] The full list of Alpha-2 country codes: List of ISO 3166 country codes - Wikipedia
[14] The full list of active currency codes: https://en.wikipedia.org/wiki/ISO_4217
[15] Browse and search for fields at ARDC – Research Vocabularies Australia: ANZSRC 2008: Fields of Research