

DReSA Automated Ingestion Provider's Guide

Version 1.4.0



Contents

1.	About DReSA	2
2.	What is Automated Ingestion	2
3.	Ingestion methods	3
	Matching resources	3
4.	Configuring a new source	3
5.	Method: REST API	3
6.	Method: Comma Separated Values (CSV) file	4
	Materials File	4
	Events File	5
7.	Method: iCalendar	6

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 methods

DReSA can find URLs via site maps or through static definitions.

Sitemaps provide a standard schema that help find appropriate pages within a site (such as: <https://dresa.org.au/sitemap.xml>). Whilst sitemaps provide a method by which information about resources can be found, a base level of data is still required to be able to register resources in DReSA. The required fields are as follows:

- **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

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'.

4. 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'

5. Method: REST API

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

The [Zenodo](#) open-access repository for ingestion of materials.

- This API provides for ingested of the following optional fields:
 - Authors - mapped from 'creators'.
 - Contributors.

The [ElixirTeSS](#) training platform for ingestion or events.

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 \(restfulapi.net\)](https://restfulapi.net)

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 ¹	Y	
URL	URL	Y	
Description	""""Text""""	Y	Can include markdown ² or html. Note ³
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 ⁴
Status	String	Y	One of the following keys: active development archived
<i>Optional Fields – the following columns can be included, in any order, or omitted completely.</i>			
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 ⁵
Audiences	String;String;		One or more semi-colon separated audience keys ⁶
Types	String;String;		One or more semi-colon separated resource type keys ⁷
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.

¹ A string is text up to a maximum of 255 characters.

² See following for reference: <https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>

³ Note: double-quote characters within the text should be replaced with two double-quotes ("").

⁴ Licence keys can be found in the following file:

https://github.com/dresa-org-au/TeSS/blob/master/config/dictionaries/target_audience.yml

⁵ Browse and search for fields at ARDC – Research Vocabularies Australia: [ANZSRC 2008: Fields of Research](#)

⁶ Target audience keys can be found in the following file:

https://github.com/dresa-org-au/TeSS/blob/master/config/dictionaries/target_audience.yml

⁷ 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
<i>Title</i>	String ⁸	Y	
<i>URL</i>	URL	Y	
<i>Description</i>	""""Text""""	Y	Can include markdown ⁹ or html. Note ¹⁰
<i>Start</i>	DateTime	Y	String in format: YYYYmmddTHHMMSS
<i>End</i>	DateTime	Y	String in format: YYYYmmddTHHMMSS
<i>Timezone</i>	String	Y	A valid TZ database name ¹¹
<i>Contact</i>	String	Y	Defaults to the Provider's contact if blank
<i>Organizer</i>	String	Y	
<i>Eligibility</i>	String;String;	Y	One or more of the following keys: open_to_all host_institution expression_of_interest by_invitation
<i>Host Institutions</i>	String;String;	Y	One or more semi-colon separated strings
<i>Online</i>	String	Y	true or false
<i>Optional Fields – the following columns can be included, in any order, or omitted completely.</i>			
<i>City</i>	String	*	Required if not online
<i>Country</i>	String	*	Required if not online. Alpha 2 code ¹²
<i>Venue</i>	""""Text""""		
<i>Postcode</i>	String		
<i>Subtitle</i>	String		
<i>Duration</i>	String		
<i>Recognition</i>	String		
<i>Types</i>	String;String;		One or more of the following keys: webinar, workshop, conference, meeting, hackathon, dropin
<i>Cost Basis</i>	String		One of the following keys: free hosts charge
<i>Currency</i>	String	*	Required if Cost Basis is not free. Three letter alphabetic currency code ¹³
<i>Cost</i>	Number	*	Required if Cost Basis is not free. Decimal value
<i>Capacity</i>	Number		Integer greater than zero
<i>Fields</i>	String;String;		One or more semi-colon separated strings ¹⁴
<i>Keywords</i>	String;String;		One or more semi-colon separated strings
<i>Audiences</i>	String;String;		One or more semi-colon separated strings
<i>Objectives</i>	""""Text""""		Can include markdown or html.
<i>Prerequisites</i>	""""Text""""		Can include markdown or html.
<i>Requirements</i>	""""Text""""		Can include markdown or html.

⁸ A string is text up to a maximum of 255 characters.

⁹ See following for reference: <https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>

¹⁰ Note: double-quote characters within the text should be replaced with two double-quotes ("").

¹¹ The full list of TZ database names: https://en.wikipedia.org/wiki/List_of_tz_database_time_zones

¹² The full list of Alpha-2 country codes: [List of ISO 3166 country codes - Wikipedia](https://en.wikipedia.org/wiki/List_of_ISO_3166_country_codes)

¹³ The full list of active currency codes: https://en.wikipedia.org/wiki/ISO_4217

¹⁴ Browse and search for fields at ARDC – Research Vocabularies Australia: [ANZSRC 2008: Fields of Research](https://www.ardc.gov.au/research-vocabularies/anzsrc-2008-fields-of-research)

7. 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 ¹⁵.

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>

¹⁵ RFC 5545: Internet Calendaring and Scheduling Core Object Specification (iCalendar).
Link as follows: <https://datatracker.ietf.org/doc/html/rfc5545>