DReSA Automated Ingestion Provider's Guide

Version 1.3.5

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 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' (materials)`

5. Method: REST API

The 'rest' method has currently only been implemented for **materials** ingestion from the Zenodo API. This API provides more fields that can be ingested into DReSA, including:

- authors, mapped from 'creators'
- contributors.

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)

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.

Events File

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

	header	format	req?	notes
1	Title	String [1]	Y	
2	URL	URL	Y	
3	Description	"""Text"""	Y	Can include markdown[2] or html
4	Start	DateTime	Y	String in format: YYYYmmddTHHMMSS
5	End	DateTime	Y	String in format: YYYYmmddTHHMMSS
6	Timezone	TZ database name [3]	Y	
7	Contact	String	Y	Default to Provider's contact if missing
8	Organizer	String	Y	

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

10	Host Institutions	String; String;	Y	One or more strings
11	Online	`true` or `false`	Y	
12	City	String	Y *	Required if not online
13	Country	Alpha 2 code [4]	Y *	Required if not online
14	Venue	String		

Materials File

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

	header	format	req?	notes
1	Title	String [1]	Y	
2	URL	URL	Y	
3	Description	"""Text"""	Y	Can include markdown or html
4	Keywords	String; String;	Y	One or more strings
5	Contact	String	Y	Default to Provider's contact if missing
6	Licence	Licence key [5]	Y	

7	Status	Status key	Y	one of the following keys:
				`active`
				`development`
				`archived`

8	Authors	String; String;	N	One or more authors
9	Contributors	String; String;	N	One or more contributors
10	DOI	String	N	

[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] The full list of TZ database names: https://en.wikipedia.org/wiki/List_of_tz_database_time_zones
[4] The full list of Alph2-2 country codes: List of ISO 3166 country codes - Wikipedia
[5] Licence keys can be found in the following file: TeSS/licences_dresa.yml at master · nrmay/TeSS · GitHub

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 [6i].

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

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