

EUROPEAN LANGUAGE GRID



Registering services in the ELG catalogue

Ian Roberts (University of Sheffield)

15/16/17-11-2021 META-FORUM 2021 – Using the European Language Grid (virtual conference)

<http://www.european-language-grid.eu>

Become a provider

- If you haven't already, you will need to register as a “provider” on the ELG
 - Sign up & sign in
 - Click “become a provider” on the <https://live.european-language-grid.eu> homepage
 - Wait for your request to be approved

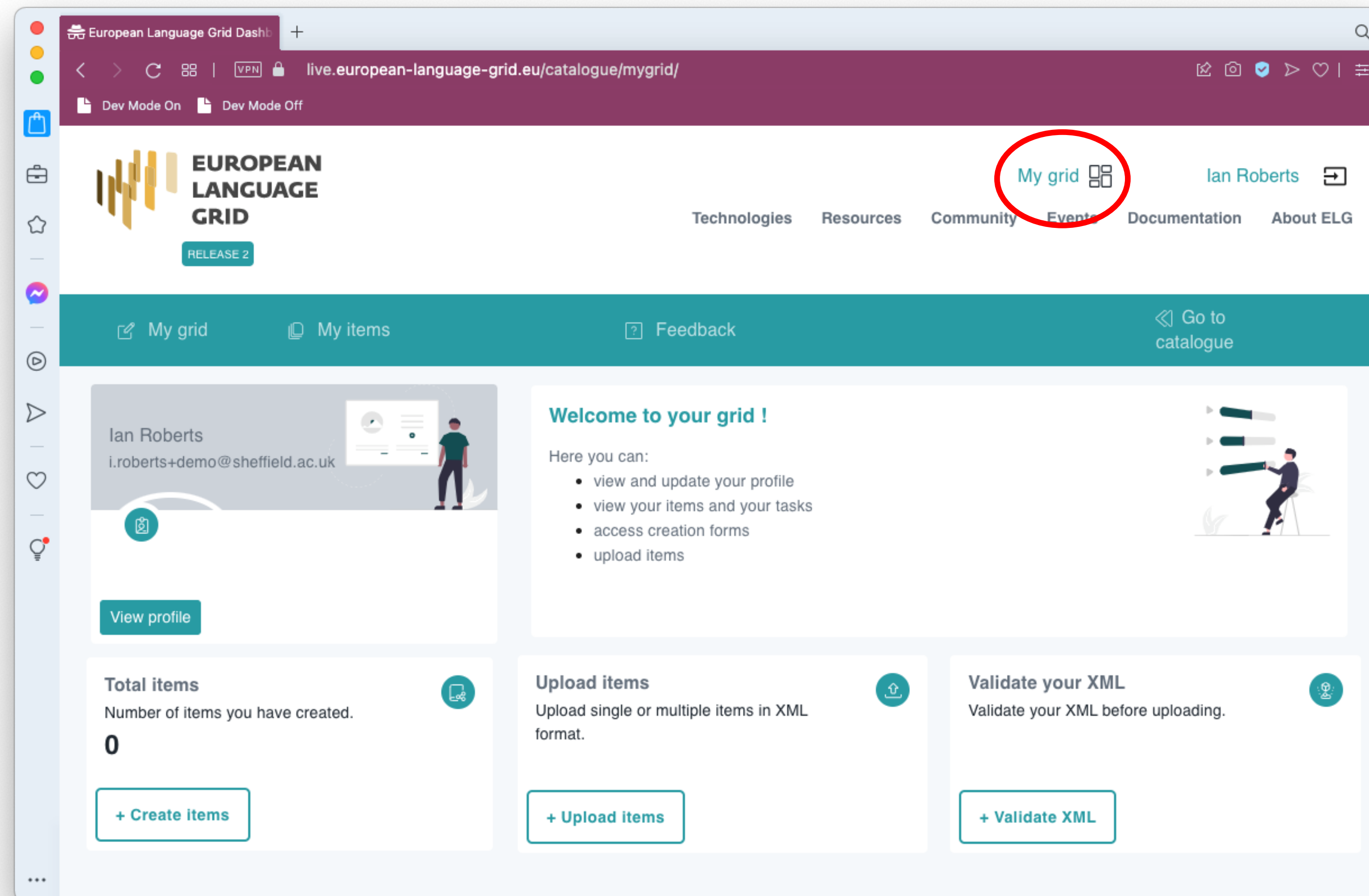
Have an amazing language resource?

ELG is primary platform for Language Technology in Europe. Your data resources, tools and technologies can be shared here and meet their users. Become a provider, and you will get acknowledgment in academic circles, increase your market, and improve the recognition of your resource.

Become a provider

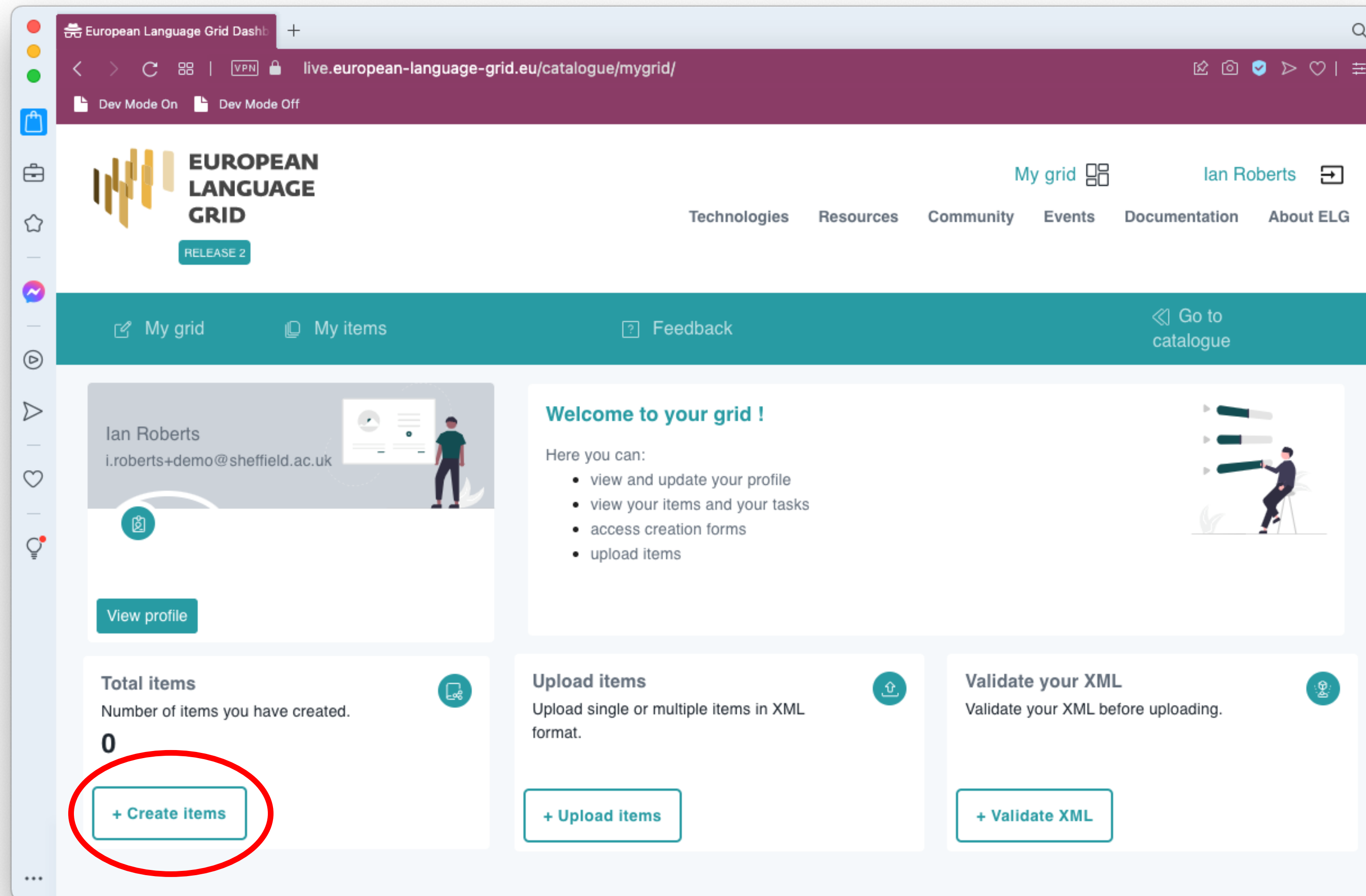
Become a provider

- Once approved as a provider, you will see “My Grid” link



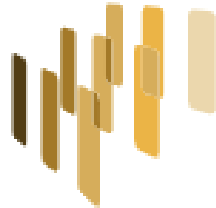
Registering your service

- Simplest approach for a small number of items is to use the metadata editor




Registering your service


- Simplest approach for a small number of items is to use the metadata editor



EUROPEAN
LANGUAGE
GRID

RELEASE 2

My grid 

Ian Roberts 

Technologies


Resources


Community


Events


Documentation

About ELG


 My grid

 My items

 Feedback


 Go to catalogue

ADD ITEMS




Service or Tool

Services that run in the cloud, downloadable tools, source code, etc., that perform language processing and/or any Language Technology related operation, such as Machine Translation, Information extraction, linguistic annotation, automatic speech recognition, etc.



Corpus

Structured collections of pieces of data, such as collections of text documents, audio transcripts, audio and video recordings, parallel corpora, linguistically annotated corpora, treebanks, etc.




Language description

Machine learning models, embeddings, language models, computational grammars

Go to form

Go to form

Go to form

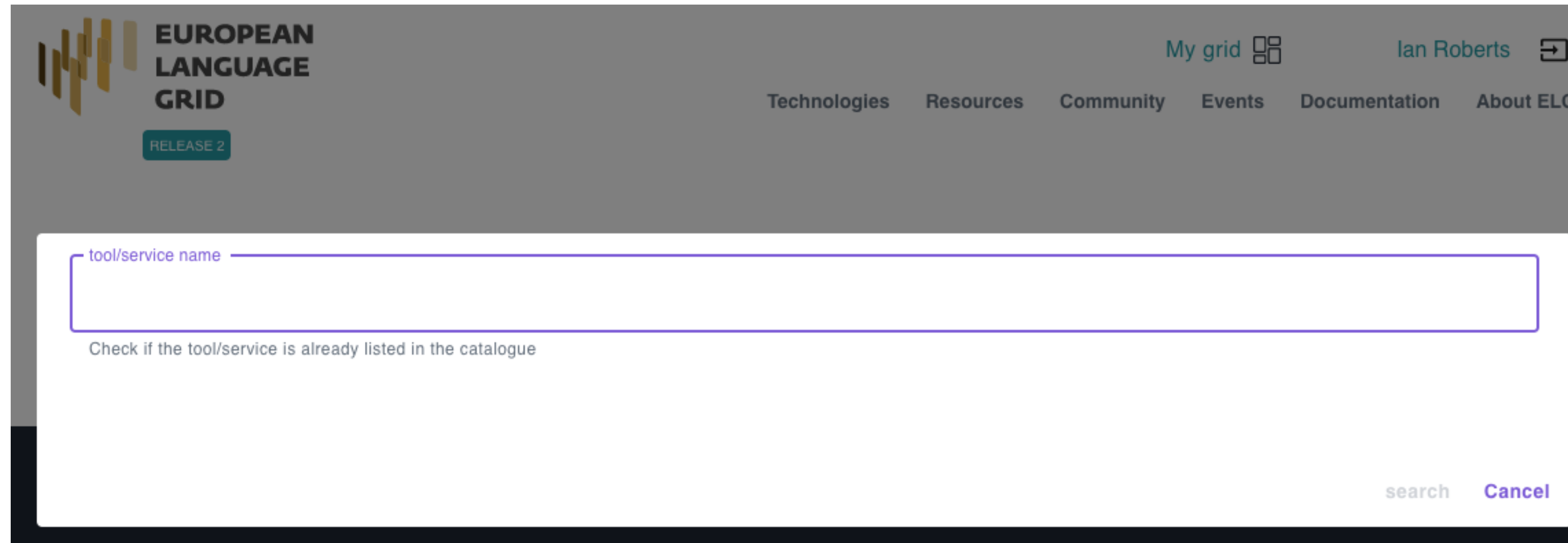


META-FORUM 2021

5

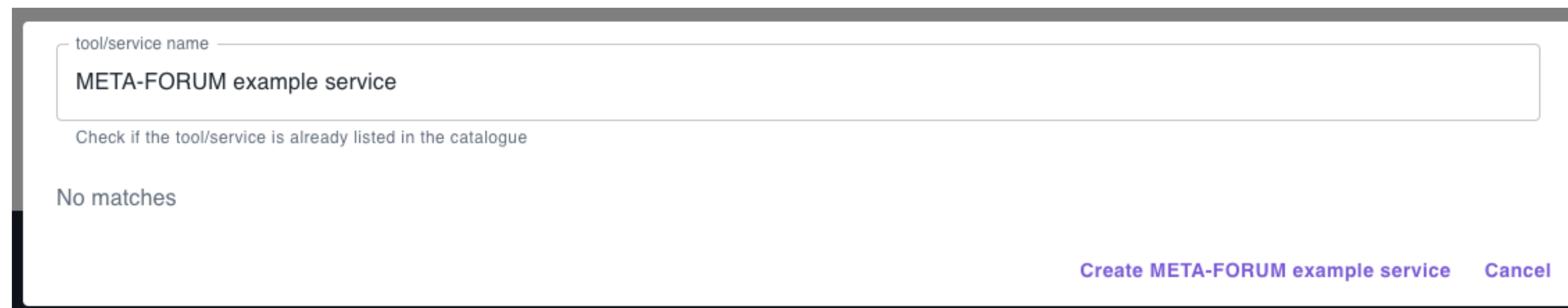
Step 1: check for duplicates

- Start by entering a name for your service, and do an initial check that it hasn't already been registered (by you or someone else!)



The screenshot shows the top navigation bar of the European Language Grid (ELG) website. The logo "EUROPEAN LANGUAGE GRID" is on the left, with a "RELEASE 2" badge below it. The navigation menu includes "Technologies", "Resources", "Community", "Events", "Documentation", and "About ELG". On the right, there are links for "My grid" and "Ian Roberts".

The main form area has a text input field labeled "tool/service name". Below the input field, it says "Check if the tool/service is already listed in the catalogue". At the bottom right of the form, there are "search" and "Cancel" buttons.



This screenshot shows the same form as the previous one, but with the text "META-FORUM example service" entered into the "tool/service name" field. The text "Check if the tool/service is already listed in the catalogue" is still present.

Below the input field, the text "No matches" is displayed. At the bottom right, the buttons are "Create META-FORUM example service" and "Cancel".

Step 2: confirm this is an ELG-compatible service

- ELG can list both tools/services that are ELG-compatible (hosted in the ELG cluster) and those that are not (e.g. links to software you can download and run locally)
- We are submitting an ELG-compatible service so we must select “yes”

Do you want to contribute a tool that will be integrated in ELG as an ELG-compatible service (i.e., available through the ELG APIs)?

☒ Yes

☐ No

You will find more information at: https://european-language-grid.readthedocs.io/en/stable/all/3_Contributing/Service.html

Step 3: Basic information common to all resources

- First tab “Language resource/technology” is for general metadata
 - Identity: **name**, **description**, **version**, logo, links to provider and funding project
 - (key required elements in **red**)

LANGUAGE
RESOURCE/TECHNOLOGY

TOOL/SERVICE

DISTRIBUTION

DATA

☐ Work in progress
☒ ELG-compatible service

Save draft

Save

IDENTITY

CATEGORIES

CONTACT

DOCUMENTATION

RELATED LRTS

LRT name *

META-FORUM example service

The official name or title of the language resource/technology

language

English

select language

LRT identifier

A string used to uniquely identify the language resource/technology

LRT short name

An abbreviation, acronym, etc. used for the language resource/technology

language

English

select language

Description

Paragraph ▼ **B** *I* U 🔗 ↶ ↷ ⚡= ⚡= <> 📄 ▼

language

English

select language

Version

1.0.0 (automatically assigned)

Recommended format: major_version.minor_version.patch (see semantic versioning guidelines at <http://semver.org>)

Version date

The date of the LRT version (latest update of the particular version if possible)

LRT provider

The actor responsible for providing, curating, maintaining and making available (publishing) the language resource/technology

Actor type

Select type and describe

Source of metadata record

The entity (repository, catalogue, archive, etc.) from which the metadata record has been imported into the new catalogue

LRT creator

The actor who created the language resource/technology

Actor type

Select type and describe

Step 3: Basic information common to all resources

- First tab “Language resource/technology” is for general metadata
 - Identity: **name**, **description**, **version**, logo, links to provider and funding project
 - Categories: **intended application**, **keywords**

The screenshot shows a web form for creating a language resource/technology. At the top, there are four tabs: 'LANGUAGE RESOURCE/TECHNOLOGY' (active), 'TOOL/SERVICE', 'DISTRIBUTION', and 'DATA'. To the right of the tabs are two checkboxes: 'Work in progress' (unchecked) and 'ELG-compatible service' (checked). Further right are two buttons: 'Save draft' and 'Save'. Below the tabs is a vertical sidebar with icons and labels for 'IDENTITY', 'CATEGORIES', 'CONTACT', 'DOCUMENTATION', and 'RELATED LRTS'. The main content area is divided into three sections: 'Intended application' with a text input field and a description; 'Domain' with a text input field and a description; and 'Keyword' with a text input field, a language dropdown menu (set to 'English'), and a plus button for adding more keywords.

LANGUAGE RESOURCE/TECHNOLOGY | TOOL/SERVICE | DISTRIBUTION | DATA

☐ Work in progress
☒ ELG-compatible service

[Save draft](#) [Save](#)

IDENTITY

CATEGORIES

CONTACT

DOCUMENTATION

RELATED LRTS

Intended application

Intended application

Select the broad LT application for which the language resource/technology is intended or recommended for (multiple values possible)

Domain

The field of knowledge or activity in which the LRT is classified (multiple values possible)

Domain label

A word or phrase describing the domain

Keyword

A word or phrase characteristic of the language resource/technology that can be used at search (multiple values possible)

Keyword *

language English

+

Step 3: Basic information common to all resources

- First tab “Language resource/technology” is for general metadata
 - Identity: **name, description, version**, logo, links to provider and funding project
 - Categories: **intended application, keywords**
 - Contact: **contact details** for support

The screenshot shows a web form for creating a language resource/technology. The top navigation bar includes tabs: LANGUAGE RESOURCE/TECHNOLOGY (active), TOOL/SERVICE, DISTRIBUTION, and DATA. On the right, there are checkboxes for 'Work in progress' (unchecked) and 'ELG-compatible service' (checked), along with 'Save draft' and 'Save' buttons. The left sidebar contains icons and labels for IDENTITY, CATEGORIES, CONTACT, DOCUMENTATION, and RELATED LRTS. The main content area is divided into sections: 'Additional information' with a description and radio buttons for 'landing_page' (selected) and 'email'; a text input field for 'Landing page *'; 'Contact' with a description and a dropdown menu for 'Actor type'; and a section for 'Documentation'.

Step 3: Basic information common to all resources

- First tab “Language resource/technology” is for general metadata
 - Identity: **name, description, version**, logo, links to provider and funding project
 - Categories: **intended application, keywords**
 - Contact: **contact details** for support
 - Documentation: links to documentation and citation that users should reference

The screenshot shows a web form for creating a language resource. The interface has a top navigation bar with tabs: 'LANGUAGE RESOURCE/TECHNOLOGY' (active), 'TOOL/SERVICE', 'DISTRIBUTION', and 'DATA'. To the right of these tabs are checkboxes for 'Work in progress' (unchecked) and 'ELG-compatible service' (checked), along with 'Save draft' and 'Save' buttons. A left sidebar contains icons and labels for 'IDENTITY', 'CATEGORIES', 'CONTACT', 'DOCUMENTATION' (highlighted in purple), and 'RELATED LRIS'. The main content area is divided into two sections: 'Documented in' and 'Preferred citation'. The 'Documented in' section has a 'Title' input field and a description: 'A document related to the language resource/technology (e.g., research paper describing its contents, its use in a project, etc.)'. The 'Preferred citation' section also has a 'Title' input field and a description: 'Add the publication you would like to be used when users cite the language resource'.

Step 3: Basic information common to all resources

- First tab “Language resource/technology” is for general metadata
 - Identity: **name, description, version**, logo, links to provider and funding project
 - Categories: **intended application**, keywords
 - Contact: **contact details** for support
 - Documentation: links to documentation and citation that users should reference
 - Related LRTs: other items that are related to this one in some way

The screenshot shows a web form for creating or editing a language resource/technology. The interface has a top navigation bar with tabs: 'LANGUAGE RESOURCE/TECHNOLOGY' (active), 'TOOL/SERVICE', 'DISTRIBUTION', and 'DATA'. To the right of the tabs are checkboxes for 'Work in progress' (unchecked) and 'ELG-compatible service' (checked), along with 'Save draft' and 'Save' buttons. A left sidebar contains icons and labels for 'IDENTITY', 'CATEGORIES', 'CONTACT', 'DOCUMENTATION', and 'RELATED LRTS'. The main content area is divided into sections: 'Previous versions' (with a description and an 'LRT name' input field), 'Version of' (with a description and an 'LRT name' input field), and 'Part of' (with a description and an 'LRT name' input field). Each input field is followed by a small text label: 'The official name or title of the resource'.

Step 4: Information specific to services

- Tool/Service tab is for all services (ELG-compatible or not)
 - Categories: service-specific categorization info, especially **function** of the service

The screenshot shows a web form for registering a service. At the top, there are tabs: LANGUAGE RESOURCE/TECHNOLOGY, TOOL/SERVICE (selected), DISTRIBUTION, and DATA. To the right of these tabs are checkboxes for 'Work in progress' (unchecked) and 'ELG-compatible service' (checked), along with 'Save draft' and 'Save' buttons. On the left side of the form, there is a vertical sidebar with three categories: CATEGORIES (selected), TECHNICAL, and EVALUATION. The main content area is titled 'Function' and contains three input fields: 1. 'Function *' with a placeholder text 'Start typing to select the task performed by the tool or add a new value (multiple values possible)'. 2. 'Wrapping framework' with a dropdown arrow and a placeholder text 'The implementation framework used to develop and run the tool/service'. 3. 'Development framework' with a placeholder text 'Start typing to select the framework or toolkit (Machine Learning model, NLP toolkit) used in the development of the tool/service'. Below this is an 'Implementation language' field with a placeholder text 'The programming language used for development of the tool'.

Step 4: Information specific to services

- Tool/Service tab is for all services (ELG-compatible or not)
 - Categories: service-specific categorization info, especially **function** of the service
- Technical: details about **input**, **output** and parameters

LANGUAGE RESOURCE/TECHNOLOGY **TOOL/SERVICE** DISTRIBUTION DATA

☐ Work in progress
☒ ELG-compatible service

Save draft Save

CATEGORIES

TECHNICAL

EVALUATION

Specify if the tool is language dependent *

☒ Language dependent

Input content resource
Describe the requirements that a data resource must fulfill in order to be processed

Input resource type *

Select the type of the resource that can be processed by the tool (e.g., corpus, single file, lexicon, etc.)

Language
The language supported for processing

Language *

Start typing to select the language

Media type

Select the media type of the data resource that can be processed by the tool

Data format

Data format

Indicates the format(s) of a data resource

Annotation type

Annotation type

If the tool takes as input an annotated resource (e.g. a parser that takes as input a tagged corpus), select the annotation type (multiple values possible)

Sample
Select to add sample texts or files that can be used for feeding the processing service for testing purposes

Fill in

Step 4: More on the “technical” tab

- If your service processes specific languages, set “language-dependent”
- Specify the language(s) on input and output resources – BCP47 tags allowing variant, script, etc.
- For text/audio analysis services (where the output is annotations rather than language), use output resource type “file”, data format “JSON”, the same language(s) as input, and select relevant annotation types
- For MT specify source languages on input resource and target languages on output resource
- If possible, supply at least one known-good “sample” for each input language
 - Sample “tag” should be a meaningful description of this sample – it will be shown in the picklist
 - Either inline “sample text” *or* upload a file as “samples location”, not both
- Parameter declarations in metadata are for the benefit of “try out” GUI – you must still verify them in code and return errors for invalid values
 - For technical reasons, the parameter names “version”, “split” and “audioOnly” are reserved

Step 4: Information specific to services

- Tool/Service tab is for all services (ELG-compatible or not)
 - Categories: service-specific categorization info, especially **function** of the service
 - Technical: details about **input**, **output** and parameters
 - Evaluation: TRL from 1 to 9

The screenshot shows the 'TOOL/SERVICE' tab selected in the top navigation bar. The left sidebar contains three main sections: 'CATEGORIES' (with a grid icon), 'TECHNICAL' (with a gear icon), and 'EVALUATION' (with a document icon). The 'EVALUATION' section is currently active. In the top right corner, there are checkboxes for 'Work in progress' (unchecked) and 'ELG-compatible service' (checked), along with 'Save draft' and 'Save' buttons. The main content area has a toggle switch labeled 'Evaluated' which is currently turned off. Below this is a dropdown menu labeled 'TRL' with a downward arrow. A small text note below the dropdown reads: 'The Technology Readiness Level of the tool; see here'. The bottom of the sidebar has a purple bar with the 'EVALUATION' label.

Step 5: Where can ELG find the code?

- Distribution tab describes Docker image
 - Distribution form: always “docker image”
 - **Private**: should we advertise the “docker pull” image location publicly?
 - **Docker download location**: image reference
[registry/]name:tag from where we can pull your image – this is *not* a URL. Please avoid “latest” tag
 - Service adapter download location: same for the adapter, if this is a main+adapter integration type
 - **Execution location**: URL of the ELG API endpoint *within the running container*, expressed as http://localhost[:port]/[path]
 - Additional h/w requirements: e.g. if your service needs lots of memory, GPU*, etc.
 - **Licence**: use existing SPDX options if possible. For closed-source services you can add your own licence terms

The screenshot shows the 'Software distribution 1' form in the ELG interface. The top navigation bar includes tabs for 'LANGUAGE RESOURCE/TECHNOLOGY', 'TOOL/SERVICE', 'DISTRIBUTION' (which is active), and 'DATA'. There are also checkboxes for 'Work in progress' and 'ELG-compatible service', and buttons for 'Save draft' and 'Save'. The form itself is titled 'Software distribution 1' and has a subtitle 'Describe separately each distributable form of the tool (e.g., docker image, source code, etc.) with information on its licensing terms'. The form contains several sections: 'Software distribution form' with a dropdown menu set to 'docker image'; 'Private' with radio buttons for 'Yes' and 'No'; 'Docker download location' with a text input field; 'Service adapter download location' with a text input field; 'Execution location' with a text input field; 'Additional h/w requirements' with a text input field; 'Documentation' with a 'Title' input field; 'Licence' with a 'Licence name' input field; 'Cost' with a 'Fill in' button; and 'Access rights' with a 'Fill in' button. At the bottom, there is a dropdown for 'For members of' and a note about selecting members if conditions or cost are reserved.

Notes

- If possible, always try to re-use existing values rather than creating your own
 - Languages, SPDX licences, annotation types, etc.
- Similarly, please create metadata records for your organization, funding projects, etc. rather than just specifying free text names
 - This helps promote your organization/project and makes it easier to find your records – ELG creates back-links from a project to all the services/resources it funded
 - Create organization first, then project linked to organization, then services linked to project
- If you have many similar records, it may be easier to
 - Create the first one through the editor
 - Export as XML, and use that as a template for others
 - Use “upload items” in “My Grid” to import them
 - We recommend going through the manual cycle at least once before switching to XML

And finally...

- Yes, the form is complex, because it needs to cover many different scenarios
- You will make mistakes!
- ... but that's OK, we are here to help
- If anything is unclear, ask us
 - details under the “feedback” link at the top of “My Grid”
 - or fill in what you can, submit the record for publication, your validator will be in touch with one-to-one assistance