EnrichEuropeana+ Integration with National Aggregator Infrastructures

Version 1.0

Documentation Information

Action Number	2020-EU-IA-0075
Project Website	https://pro.europeana.eu/project/enricheuropeana
Contractual Deadline	30/10/2022
Nature	Description of work
Author	Marcin Heliński
Contributors	Kathryn Cassidy, Tommy Schmucker
Reviewer	Sergiu Gordea, Tomasz Parkoła, Frank Drauschke
Version	1.0
Date	16/11/2022

Co-financed by the Connecting Europe Facility of the European Union

Contents

Introduction

EnrichEuropeana+ (fully titled ‘Enriching Europeana through citizen science and artificial intelligence - unlocking the 19th century') aims to enhance Europeana Transcribe (www.transcribathon.eu) as a service for cultural heritage institutions.

Scope

This document describes the work done for making the crowdsource enrichments available for national aggregators and content providers. By using the Data Exchange Infrastructure, certain metadata enrichments created by the Transcribathon users within the crowdsourcing campaigns are exposed for validation in the national aggregator or content provider infrastructures. The workflow of the enrichments and also the specification of API requests engaged in this process are described in the following sections. Validation of enrichments requires involvement of professional users both on national aggregator and content provider level. Therefore, this document describes changes implemented into the GUI of the digital collections management software used within the aforementioned institutions. Conceptual work and software development have been done so far which provide means to introduce enrichment workflow as envisaged by the project. However, during the analysis and investigation of various aspects related to enrichment workflow, several improvements have been identified. Hence, there is a dedicated section of this document describing possible next steps to make the enrichment workflow even more advanced and complete.

Objectives

The main objectives of EnrichEuropeana+ are:

- To engage public users and professionals in enhancing the semantic and multilingual description of Cultural Heritage objects by continuing the development of Europeana Transcribe.
- To increase accessibility of manuscripts related to historical events and societal transformations in Europe within the 19th Century through a new Citizen Science crowdsourcing campaign to stimulate user engagement for transcribing, translating, and adding semantic enrichments.
- To transform Europeana Transcribe into a service used by Cultural Heritage Institutions to crowdsource the enrichment of cultural object descriptions and improve the multilingualism of metadata.

The main objective of the *Integration with National Aggregator Infrastructures* milestone is to enable national aggregation infrastructure to benefit from the enrichments created in the Transcribathon platform. This has been done via a feedback loop feature implemented in the Data Exchange Infrastructure that notifies content providers (or national aggregators if they act in this role) about newly available enrichments, giving them an opportunity to accept or reject those enrichments in their internal validation procedure (at the content provider side). Once the data are accepted, they become part of the original object's metadata and can be transferred into the Europeana platform via national aggregation infrastructure, using existing aggregation routines.

Metadata enrichment scenarios

One of the main goals of this project was integration of enrichments collected from the Transcribathon Platform with the original records held in the content providers' or national aggregators' systems. While designing this process, the project partners identified the contextual classes and attributes considered most valuable and practically applicable to integrate crowdsourced enrichments. The dates and places have been the attributes of the first choice. Both attributes may be used for improving the quality of the record in Europeana as the provided values can be converted to contextual classes edm:TimeSpan and edm:Place in the EDM record[1].

Date scenario

The current state of the Transcribathon user interface allows one to specify a general date, but it does not allow indicating the exact meaning of the created date enrichments. Therefore the first approach to date enrichments assumes placing the user created values in the general Dublin Core attribute *dc:date*. Although such enrichment is not the most precise, it still significantly improves the quality of the record in Europeana. Transferring dates to other attributes is planned for the final release of the software developed in this project, however, a set of possible options need to be agreed upon based on typical actions executed along the Transcribathon events.

Place scenario

The current state of the Transcribathon user interface allows one to specify places related to the transcribed content. The places mentioned in the records content, being under transcription, will be transferred to content providers or national aggregators and included in the *dcterms:spatial* attribute of the original record. Information provided by the users will be integrated in the original record in a different way on the content provider and on the national aggregator side. Nethertheless the final result in the EDM record will be a new contextual class *edm:Place* containing the provided information.

[1] Metadata Enrichment Scenarios is a result of analysis made by the partners and described in the documents https://docs.google.com/document/d/1AOMA4ufxOXn8lM_JM6lqn-n-Tiakkrtq2nPYK5gXMqo/edit and https://docs.google.com/document/d/133TNlhrRTd70cFfZ6-UYhDV4qmAAO_hmuFy8aggHEuo/edit

Metadata enrichment workflow

There are several steps or phases in the metadata enrichment workflow. First, users create enrichments in the Transcribathon Platform. Once the information is validated by the privileged user, it is transferred to the Data Exchange Infrastructure. At this point, Data Exchange Infrastructure informs content providers or national aggregation platforms (if it acts as a content provider) about the availability of these new enrichments. Professionals of content provider institutions perform the final validation of enrichments and the Data Exchange Infrastructure is informed about the approval or rejection of specific enrichments. After updating the original records, content providers use the regular aggregation routine to re-ingest the updated objects (in case there was an approval, i.e. an update of an object) into the Europeana platform. Once the information is ingested into Europeana platform, the process is finished, but potentially can be again initiated, if new enrichments appear in relation to the given record. The diagram on Figure 1 shows the flow of enrichments in certain phases.

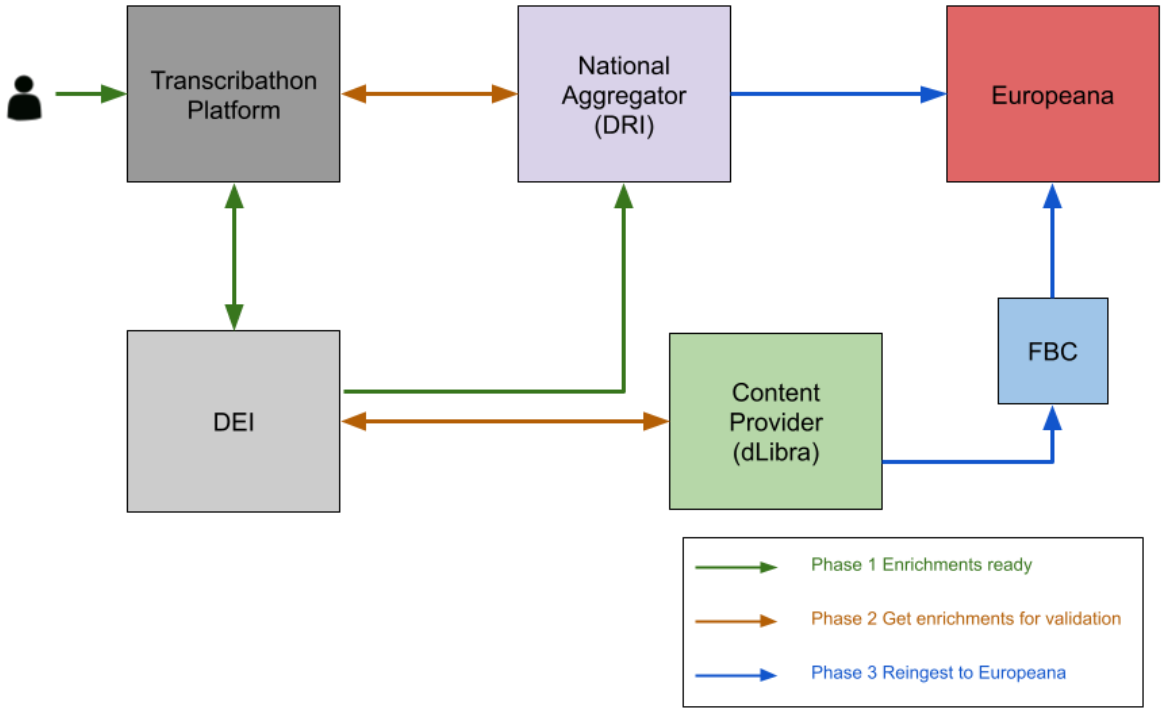

Figure 1. Metadata enrichments workflow - phases

Phase 1 - notification of available enrichments

Transcribathon Platform allows users to create transcriptions of the documents but this is not the only purpose of this system. Another important functionality includes creating metadata enrichments for the transcribed documents, for example, adding dates, places or persons mentioned in the transcribed texts. Figure 1. shows the information added by a Transcribathon user on a given document.

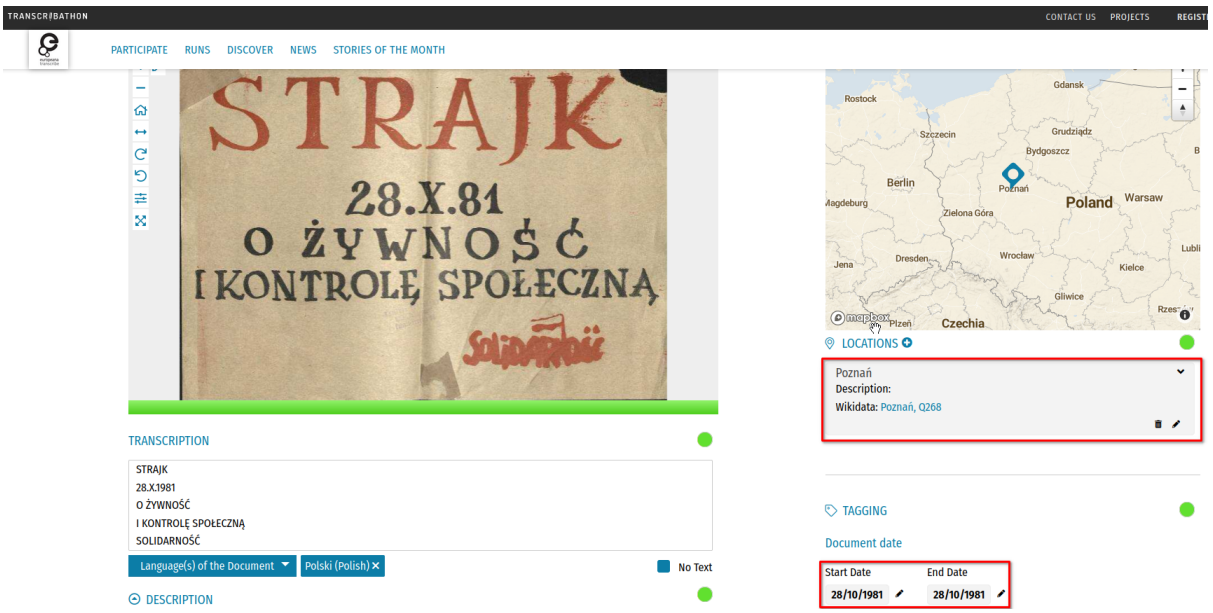

Figure 2. Transcribathon user contribution for location and dates.

The values created in the crowdsourcing process have to be validated by the privileged users within the Transcribathon tool. Once the Transcribathon documents achieve the reviewed status, Transcribathon will inform DEI about this fact by sending an API request with the record identifier. In response to that request, DEI creates an internal task which collects all available enrichments for the specified record. That is done by sending the request to Transcribathon API and processing the response to extract necessary information. The last step of the task is notifying the national aggregator and content provider about available enrichments. Within the scope of this project support for integration with the content management tool dLibra and DRI national aggregator was implemented. In case of dLibra software it is not possible to use the REST API for informing the server about possible enrichments. Therefore, dLibra software will have a mechanism for checking the availability of new enrichments for the domain of the certain dLibra instance. Informing a national aggregator, DRI in this case, will be done by sending the request to the REST API endpoint. The following diagram shows the requests flow for the part of the process described above, and the REST API requests used in this phase of the process are described in the following subsections.

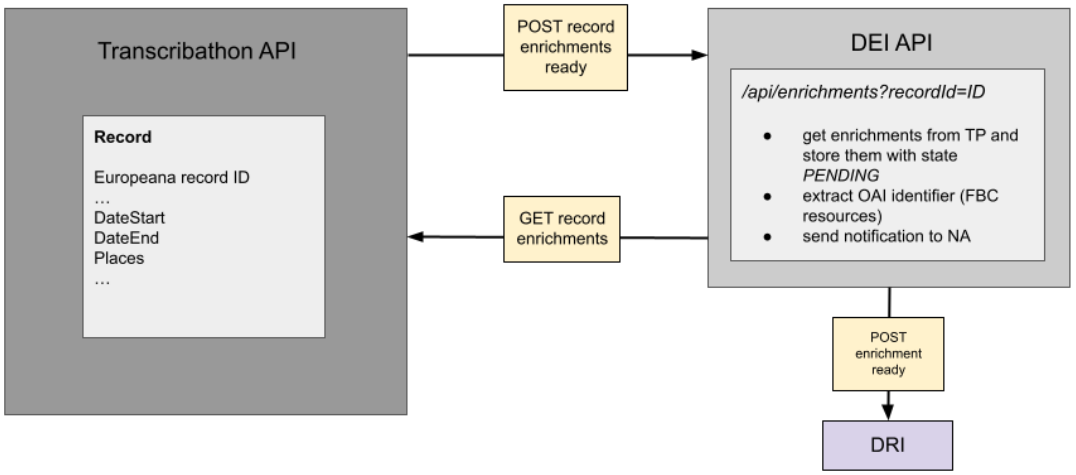

Figure 3. Request flow diagram for enrichments ready for validation

New enrichments available (DEI API)

This request is used to inform DEI that enrichments for a record identified by *recordId* are available in Transcribathon. This notification will be sent to the DEI API by Transcribathon API.

Type	*POST*
Endpoint	*/api/enrichments*
Authorization	*OAuth 2.0 Access Token*
Query Params	*recordId* - Europeana identifier of the record
Response	*200 OK* *400 Bad Request* *404 Not found*

Example

Request

https://fresenia.man.poznan.pl/dei/api/enrichments?recordId=/2020601/https__1914_1918_europeana_eu_contributions_21756
Response
No response body, 200 OK

New enrichments available (DRI API)

This request is used for informing DRI that there are enrichments available for a record identified by *recordId*. It is implemented in DRI API and is called by DEI once it has retrieved new enrichments from TP API.

Type	*POST*
Endpoint	*/enrichments*
Authorization	*Authorisation token passed as url parameter (user_email and user_token)*
Query Params	*recordId - Europeana identifier of the record sent as a url parameter*
Response	*204 No Content* *400 Bad Request* *404 Not found*

Example

Request
https://repository.dri.ie/enrichments?recordId=/719/_p841pc90p&user_email=dei@man.poznan.pl&user_token=XYZ
Response
No response body, 204 No Content

Get metadata enrichments for record (TP API)

This request is used for retrieving information about the story and its items. It is implemented in TP API and is called by clients like DEI or DRI. The response contains metadata enrichments related to the whole record (story) and the individual items. The client may extract the enrichments from the JSON response.

Type	*GET*
Endpoint	*/tp-api/stories*
Authorization	*None*
Query Params	*recordId* - Europeana identifier of the record
Response	*200 OK*

Example

Request
https://transcribathon.eu/tp-api/stories?recordId=/2025902/_nnVvzTb
Response (shortened)

```
[
        {
        "ExternalRecordId": "http://data.europeana.eu/item/135/_nnVvT3m",
        "Items": [
        {
                "Annotations": [],
                "AutomatedEnrichments": [],
                "Comments": [],
                "CompletionStatusColorCode": "#fff700",
                "CompletionStatusId": 2,
                "CompletionStatusName": "Edit",
```

```
                "DatasetId": 0,

                "DateEnd": "1988-12-31T23:00:00Z[UTC]",

                "DateEndDisplay": "1989",

                "DateStart": "1988-12-31T23:00:00Z[UTC]",

                "DateStartDisplay": "1989",

                "Description": "Uniform und Kleiderplane von Bausoldat Frank Drauschke item 1",

                "DescriptionLanguage": 5,

                "ImageLink":
"{\"@id\":\"rhus-209.man.poznan.pl/fcgi-bin/iipsrv.fcgi?IIIF=1//135/_nnVvT3m/W18_7438_crop.tif/full/
full/0/default.jpg\",\"@type\":\"dctypes:Image\",\"width\":1156,\"height\":2300,\"service\":{\"@id\"
:\"rhus-209.man.poznan.pl/fcgi-bin/iipsrv.fcgi?IIIF=1//135/_nnVvT3m/W18_7438_crop.tif\",\"@context\"
:\"http://iiif.io/api/image/2/context.json\",\"profile\":\"http://iiif.io/api/image/2/level1.json\"}
}",

                "ItemId": 434990,

                "LockedTime": "2020-04-03 17:19:47",

                "LockedUser": 74,

                "Manifest":
"https://fresenia.man.poznan.pl/dei/api/transcription/iiif/manifest?recordId=/135/_nnVvT3m",

                "OldItemId": 0,

                "OrderIndex": 1,

                "Persons": [],

                "Places": [

                {

                        "Comment": "",

                        "Latitude": 53.5,

                        "Link": "",

                        "Longitude": 14,

                        "Name": "Pasewalk",

                        "PlaceId": 68342,

                        "UserGenerated": "1",

                        "UserId": 90,

                        "WikidataId": " Q493924",

                        "WikidataName": "Pasewalk",

                        "Zoom": 10

                }

                ],

                "Properties": [],

                "Timestamp": "2019-09-23 07:26:35",

                "Title": "Uniform und Kleiderplane von Bausoldat Frank Drauschke Item 1",

                "Transcriptions": []
```

```
        }

        ],

        "OrderIndex": 0,

        "ParentStory": 0,

        "PlaceLatitude": 0,

        "PlaceLongitude": 0,

        "PlaceUserGenerated": "1",

        "PlaceUserId": 0,

        "PlaceZoom": "10",

        "PreviewImage":
"{\"@id\":\"rhus-209.man.poznan.pl/fcgi-bin/iipsrv.fcgi?IIIF=1//135/_nnVvT3m/W18_7438_crop.tif/full/
full/0/default.jpg\",\"@type\":\"dctypes:Image\",\"width\":1156,\"height\":2300,\"service\":{\"@id\"
:\"rhus-209.man.poznan.pl/fcgi-bin/iipsrv.fcgi?IIIF=1//135/_nnVvT3m/W18_7438_crop.tif\",\"@context\"
:\"http://iiif.io/api/image/2/context.json\",\"profile\":\"http://iiif.io/api/image/2/level1.json\"}
}",

        "ProjectId": 11,

        "StoryId": 12855,

        "dcCreator": "#agentOf:nnVvT3m_1",

        "dcDescription": "Zum 02. November 1989 wurde Frank Drauschke, der in Berlin-Weißensee
lebte, als Bausoldat nach Pasewalk eingezogen. Dadurch war er bei den historischen Ereignissen zur
Demonstration am 04. November auf dem Alexanderplatz und zum Fall der Mauer am 09. November nicht in
Berlin. Während seiner Grundausbildung in Pasewalk haben die Bausoldaten gemeinsam die Aktuelle
Kamera geguckt. Sie haben auch die Pressekonferenz mit Günther Schabowski am 09. November gesehen,
wussten aber nicht, was die Aussage wirklich bedeutete. Am 10. November kamen sie nach dem
Marschieren zurück und hörten erstaunt die Berichte über die Öffnung von neuen Grenzübergängen. Sie
glaubten, sie waren im falschen Film. Normalerweise musste man während der Dienstzeit seinen
Personalausweis abgeben und hatte nur noch den Wehrdienstausweis bei sich. Frank Drauschke tat dies
jedoch wie viele andere nicht, sodass er den Ausweis seinen Eltern am 12.11.1989 mitgeben konnte,
die ihm ein Visum einstempeln ließen. Am 19. November gab es den ersten Ausgang für den Ort Pasewalk
und Umgebung. Frank Drauschke wurde von Freunden abgeholt und gemeinsam fuhren sie unerlaubterweise
nach Berlin. Der erste Weg war natürlich in Zivil über den Grenzübergang Bornholmer Straße, wo Frank
Drauschke als Kind gewohnt hatte, nach Westberlin. Die Ausgänger hätten um 24 Uhr wieder in der
Kaserne zun sein. Allerdings verfuhr sich der Fahrer mehrfach auf der Autobahn, sodass sie erst um 3
Uhr nachts zurück kehrten. Die Ausweise wurden versteckt und die Zuspätkommer erzählten, sie wären
im Nachbarort auf einer Hochzeit gewesen, hätten zu viel getrunken und in die Kaserne laufen müssen.
Der Spieß akzeptierte das in gewisser Weise als Entschuldigung, den Bausoldaten wurden „nur“ drei
Wochen Ausgangs- und Urlaubssperre auferlegt.In der Folgezeit versuchten nun alle aktiv, in eine Art
Zivildienst versetzt zu werden. Der Kompaniechef der Baueinheit 15, Major Thurm, verlas ein
Fernschreiben, in dem stand, dass die Bausoldaten der Einheit am 04.12.1989 über Prenzlau ins
Gesundheitswesen versetzt werden sollten. Am 04.12. 1989 wurden 60 Bausoldaten jedoch nicht nach
Prenzlau, sondern in ein Betonwerk in Götschendorf gebracht. In der ersten Nacht (es gab ein
3-Schicht-System) streikten einige Bausoldaten und verfassten Beschwerdebriefe. Am nächsten Tag
legten sie die Beschwerden Offizieren aus Berlin vor, die sie beruhigten, es handle sich nur um eine
Übergangssituation. Darauf willigten die Soldaten ein, bezogen die überfüllten Unterkünfte und
arbeiteten zwei Wochen im 3-Schicht-System. Im Anschluss wurden zuerst die Bausoldaten, die
gestreikt hatten, fernab ihrer Heimatorte, versetzt. Frank Drauschke sollte sich im Krankenhaus
Demmin melden. Vor seiner Reise dorthin fuhr er nach Berlin und schmiss den Sack mit Uniformen auf
den Dachboden seiner Eltern, wo er 25 Jahre lang lag. Er trat seinen Dienst in Demmin an, wurde
jedoch gleich weiter nach Tutow in ein Alters- und Pflegeheim versetzt. Dort arbeitete er bis Ende
Januar als Heizer. Danach bemühte er sich erfolgreich, ins Krankenhaus Berlin-Weißensee versetzt zu
werden. Seinen Dienst dort trat er am 01.02.1990 an und blieb bis zum 31.08.1990. Zu diesem Termin
hatte er seine vorzeitige Entlassung beantragt, um ab 1. September 1990 Abitur in der EOS ''Carl von
Ossietzky'' in Berlin-Pankow machen zu können.",

        "dcLanguage": "deu",
```

```
        "dcRights": "http://creativecommons.org/licenses/by-sa/3.0/pl/ Creative Commons
Namensnennung -Weitergabe unter gleichen Bedingungen (CC-BY-SA)",

        "dcTitle": "Uniform und Kleiderplane von Bausoldat Frank Drauschke",

        "dcType": "Inny",

        "edmCountry": "Germany",

        "edmDataProvider": "Europeana 1989",

        "edmDatasetName": "135_Ag_EU_1989_Germany",

        "edmIsShownAt": "http://fbc.pionier.net.pl/id/oai:europeana1989.eu:1732",

        "edmLandingPage": "https://www.europeana.eu/portal/record/135/_nnVvT3m.html",

        "edmLanguage": "de",

        "edmProvider": "Europeana 1989",

        "edmRights": "http://creativecommons.org/licenses/by-sa/3.0/pl/"

        }

]
```


Get metadata enrichments for each item (TP API - DRI)

This request is used for retrieving the detailed information about each item in a story. DRI uses the available endpoint in the Transcribathon API for fetching available enrichments. The response contains more detail than provided by the stories API above, including all enrichments as well as the transcriptions. The client may extract the enrichments from the JSON response. This API does not provide access to any automated HTR transcriptions, and so further development is still anticipated in order to integrate these automated transcriptions with the National Aggregator and Content Provider systems.

Type	*GET*
Endpoint	*/tp-api/items/[itemId]*
Authorization	*None*
Query Params	*None*
Response	*200 OK*

Example

Request

https://transcribathon.eu/tp-api/items/39377941

Response (shortened)

{

"ItemId": 39377941,

"Title": "Plan of the stables of the Provost of Trinity College and of improvements of Nassau Street and Clare Street. Item 1",

…

"Description": "Architectural drawing of base wall and railing for Trinity College Dublin. The drawing shows two views of the proposed railing and contains numerous measurements and a scale along the bottom of the page.\n\nIt is signed by the architect Frederick Darley, and a number of Wide Street Commission members.",

"DescriptionLanguage": 7,

"DateStart": "Jul 10, 1841 12:00:00 AM",

"DateStartDisplay": "10/07/1841",

"DateEndDisplay": "",

"DatasetId": 0,

"ImageLink":
"{\"@id\":\"https://repository.dri.ie/loris/ht259422g:j099bw900/full/full/0/default.jpg\",\"@type\":\"dctypes:Image\",\"format\":\"image/jpeg\",\"width\":4568,\"height\":3288,\"service\":{\"@context\":\"http://iiif.io/api/image/2/context.json\",\"@id\":\"https://repository.dri.ie/loris/ht259422g:j099bw900\",\"profile\":\"https://iiif.io/api/image/2/level2.json\"}}",

…

"Properties": [

{

"PropertyId": 3,

"PropertyValue": "Handwritten",

"PropertyDescription": "NULL",

"PropertyTypeId": 3,

"PropertyType": "Category"

},

{

"PropertyId": 5,

"PropertyValue": "Printed",

"PropertyDescription": "NULL",

"PropertyTypeId": 3,

"PropertyType": "Category"

},

```
{
"PropertyId": 94,
"PropertyValue": "Drawing",
"PropertyDescription": "NULL",
"PropertyTypeId": 3,
"PropertyType": "Category"
}
],
"Places": [
{
"PlaceId": 81016,
"Name": "Nassau Street",
"Latitude": 53.3426,
"Longitude": -6.25702,
"ItemId": 39377941,
"Link": "",
"Zoom": 10,
"Comment": "",
"WikidataName": "Nassau Street",
"WikidataId": " Q6967582",
"UserGenerated": "1",
"UserId": 4219
}
],
"Persons": [
{
"PersonId": 9849,
"FirstName": "Frederick ",
"LastName": "Darley",
"BirthPlace": "Dublin",
"BirthDate": "NULL",
"DeathPlace": "Dún Laoghaire",
"DeathDate": "NULL",
"Link": "NULL",
"Description": "Architect; Councillor, St. Stephens Ward, Dublin City Council",
"ItemId": 39377941
},
```

```
{

"PersonId": 10228,

"FirstName": "Isaac Matthew",

"LastName": "DOlier",

"BirthPlace": "NULL",

"BirthDate": "NULL",

"DeathPlace": "NULL",

"DeathDate": "NULL",

"Link": "NULL",

"Description": "NULL",

"ItemId": 39377941

}

],

"Transcriptions": [

{

"TranscriptionId": 329451,

"Text": "TRINITY COLLEGE DUBLIN.<br /><br /><br />&lt;----- 8.0 -----&gt;  <span
class=\"pos-in-text\">(height of railings as shown in drawing on left)</span><br />&lt;----- 5.0
-----&gt; <span class=\"pos-in-text\">(height of wall as shown in drawing on left)</span>  <br
/>level of pathway<br /><br /><br />&lt;----- 2.6 -----&gt;&lt;----- 1.3 -----&gt;  X 2.6<br
/>&lt;----- 2<img class=\"tct_missing\" style=\"display: inline;\"
src=\"../../../wp-content/themes/transcribathon/images/tinyMCEImages/missing.gif\" alt=\"missing\"
/> -----&gt;&lt;-----  1.3 -----&gt; X 2.0<br />level of pathway<br />&lt;----- 2.6
-----&gt;<br />&lt;----- average depth 8.0 -----&gt;<br />&lt;----- 2.6 -----&gt;&lt;----- 2.0
-----&gt; X 2.0<br />&lt;----- 3.0 -----&gt;<br /><br />Isaac M D'Olier<br />Richard <img
class=\"tct_missing\" style=\"display: inline;\"
src=\"../../../wp-content/themes/transcribathon/images/tinyMCEImages/missing.gif\" alt=\"missing\"
/><br />Frederick Darley<br />July 10th. 1841<br /><br />Wm. Edington<br />Present <img
class=\"tct_missing\" style=\"display: inline;\"
src=\"../../../wp-content/themes/transcribathon/images/tinyMCEImages/missing.gif\" alt=\"missing\"
/><br /><br /><br /><br />WORKING DRAWING.<br />for proposed Base &amp; Iron railing.<br />TO BE
USED IN THE IMPROVEMENT OF NASSAU ST.<br /><br />Frederick Darley. Archt.<br /><br /><br /><br
/>0     1     2     3     4   
 5     6     7     8     9 
   10  <span class=\"pos-in-text\">(scale drawn at bottom of page)</span>  ",

"TextNoTags": "TRINITY COLLEGE DUBLIN.\n\n\n<----- 8.0 ----->  (height of railings as shown in
drawing on left)\n<----- 5.0 -----> (height of wall as shown in drawing on left)  \nlevel of
pathway\n\n\n<----- 2.6 -----><----- 1.3 ----->  X 2.6\n<----- 2 -----><-----  1.3 -----> X
2.0\nlevel of pathway\n<----- 2.6 -----><----- average depth 8.0 -----><----- 2.6 -----><-----
2.0 -----> X 2.0\n<----- 3.0 -----><\n\n\nIsaac M D'Olier\nRichard \nFrederick Darley\nJuly 10th.
1841\n\nWm. Edington\nPresent \n\n\n\nWORKING DRAWING.\nfor proposed Base & Iron railing.\nTO BE
USED IN THE IMPROVEMENT OF NASSAU ST.\n\nFrederick Darley. Archt.\n\n\n\n0     1     2     3     4
5     6     7     8     9     10  (scale drawn at bottom of page)  \n\n",

"Timestamp": "2022-03-29 13:21:00",

"UserId": 4219,

"WP_UserId": 3326,

"ItemId": 39377941,

"CurrentVersion": "1",
```

```
"NoText": "0",
"Languages": [
{
"LanguageId": 7,
"Name": "English",
"NameEnglish": "English",
"ShortName": "EN",
"Code": "en"
}
]
}
],
"AutomatedEnrichments": []
}
```


Phase 2 - validation of enrichments

This phase is different for content providers based on dLibra software (Europeana 1989, UWr) and national aggregators (DRI). Figure 4. shows the flow of requests in case of dLibra content providers.

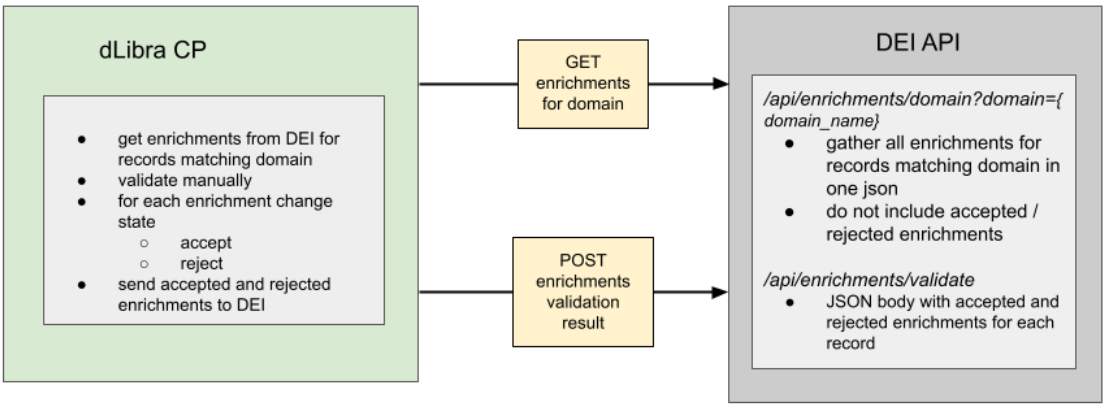

Figure 4. Enrichment validation for dLibra content providers

dLibra software contains a mechanism for checking availability of enrichments for records related to its domain. Data Exchange Infrastructure API exposes a dedicated endpoint for this. After receiving a request with a certain domain DEI gathers all the enrichments for records

matching it. By default only enrichments with PENDING state are included. The response contains JSON structure with all the enrichments. After collecting the enrichments dLibra editor users have the possibility of reviewing them. For each enrichment there is a value and other necessary information depending on the enrichment type including link to the item page in Transcribathon Platform. Item link is skipped for enrichments on the record level. dLibra collects the state for each reviewed enrichment and sends the validation result in one request to the DEI API. On the DEI side the state of the enrichments is changed according to the information from the request.

Development in dLibra software

A new module that is able to communicate with DEI has been created in dLibra Server[2]. For proper functioning, it is necessary to include in the configuration the DEI service URL, authorization data used by the dLibra server to retrieve the access token and the digital library domain that corresponds to this server. dLibra server does not expose any REST API and therefore cannot accept requests from other services. This is the main reason for implementing a mechanism of periodical polling DEI in order to check whether there are any enrichments for the records matching the digital library domain specified in the server configuration. The metadata enrichments have to be validated by the dLibra users in the Editor's Application. Once the dLibra server receives enrichments from DEI it uses the internal messaging mechanism to inform editors about this fact. Figure 5. shows the flow of requests when working with enrichments.

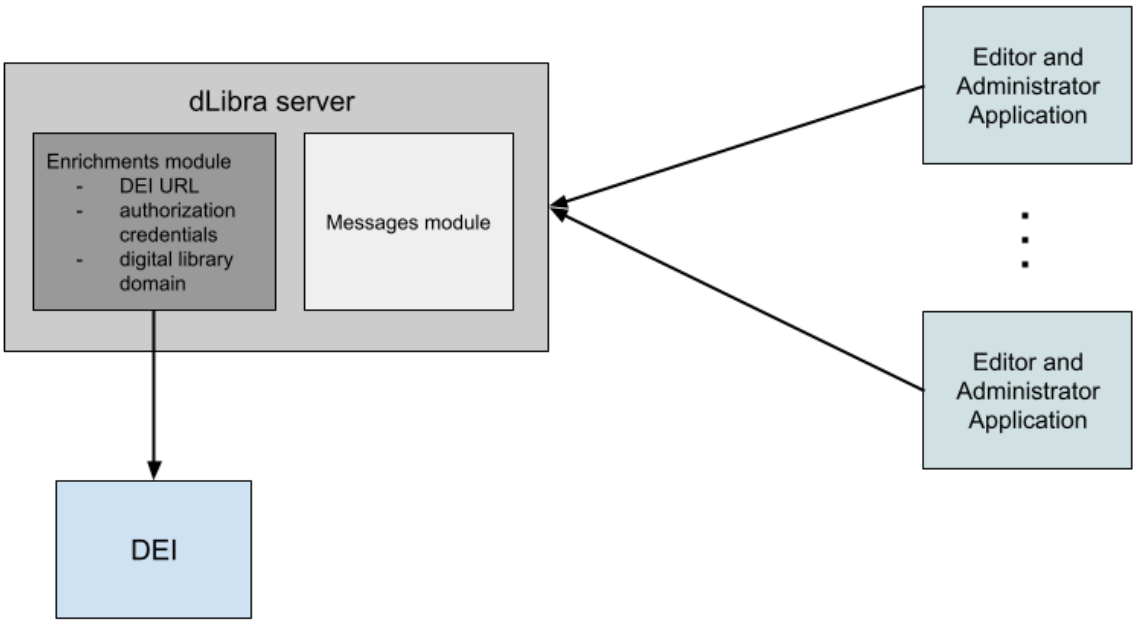

Figure 5. Communication between dLibra server, Editor and Administrator Application and DEI

[2] dLibra Server is a backend module of dLibra software, the documentation can be found at https://docs.psnc.pl/display/DLI6EN/%5BEN%5D+04.+The+dLibra+System+Server

dLibra Editor's Application[3] is a desktop application allowing users to work with digital objects. A new plugin for managing enrichments has been developed within this application. It is responsible for downloading the metadata enrichments waiting for validation in dLibra server. The enrichments of both date and place types are displayed for each record. Every enrichment entry contains information retrieved from Transcribathon via DEI. If the enrichment is associated with the certain page of the record it is possible to follow the item link to the Transcribathon Platform. For place enrichments there is also possibility to follow the link to the WikiData service if it's provided.

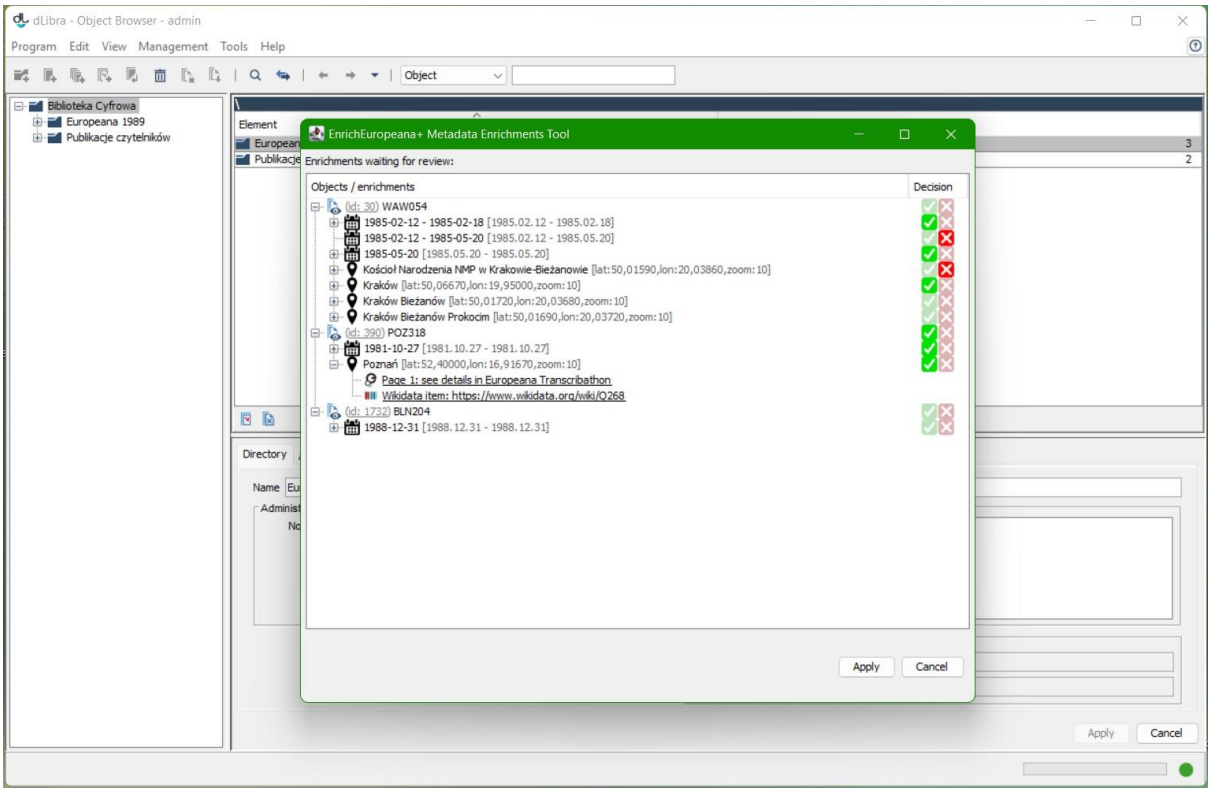

Figure 6. Metadata Enrichments Tool

For each enrichment the user has to make a decision on accepting or rejecting it. Once the decisions are made they can be applied to the enrichments stored in the server module. Before submitting the enrichment decisions a summary of this operation is displayed to the user for final confirmation. Figure 6 shows the confirmation message.

[3] The documentation can be found at https://docs.psnc.pl/display/DLI6EN/%5BEN%5D+05.+The+Editor+and+Administrator+Application

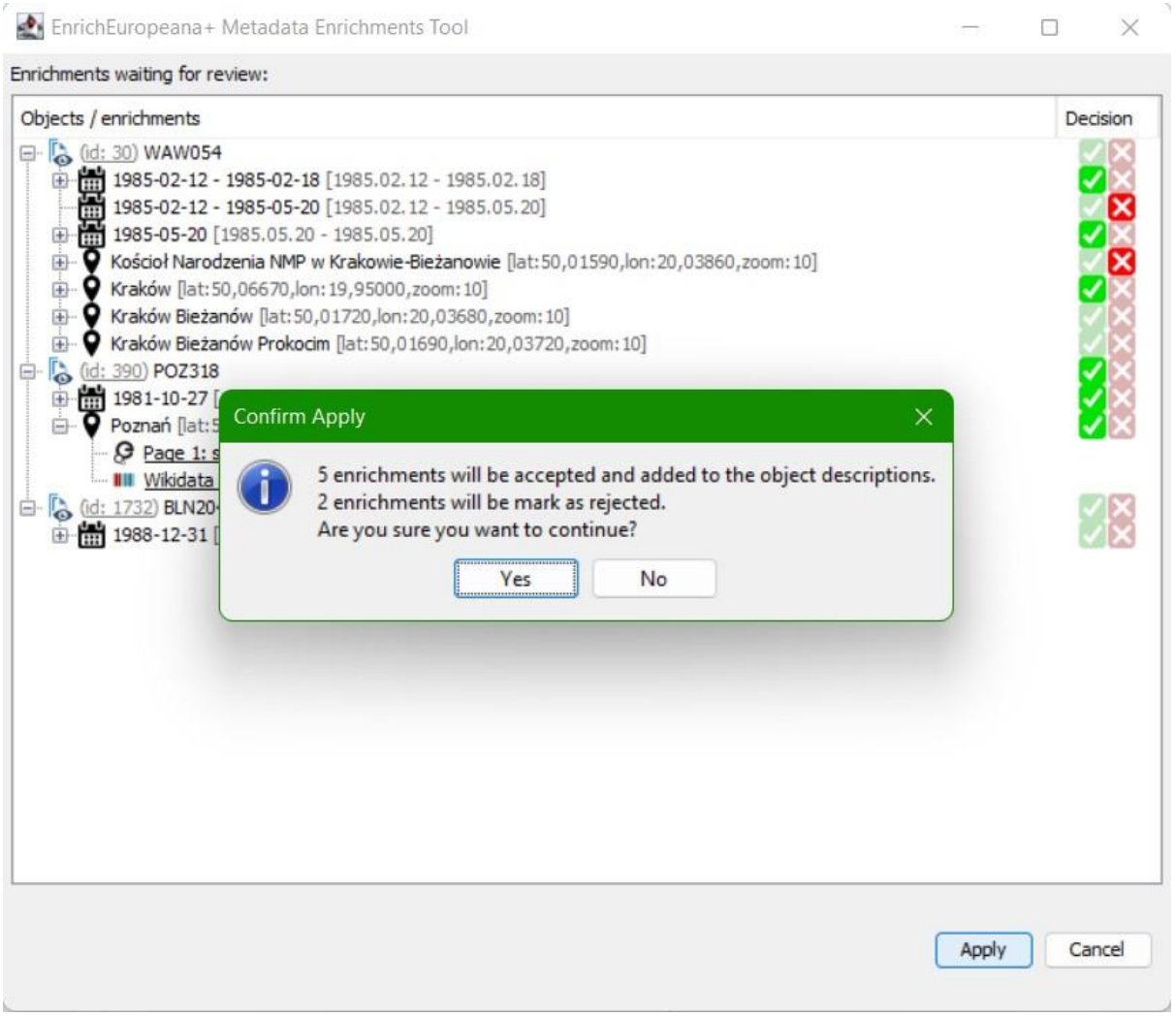

Figure 7. Confirmation message

Confirming the user's choice triggers the submission of the new status of the enrichments to the dLibra server module and further to DEI. All accepted and rejected enrichments are sent in one request. The accepted ones appear in the original record metadata like shown in the Figure 8.

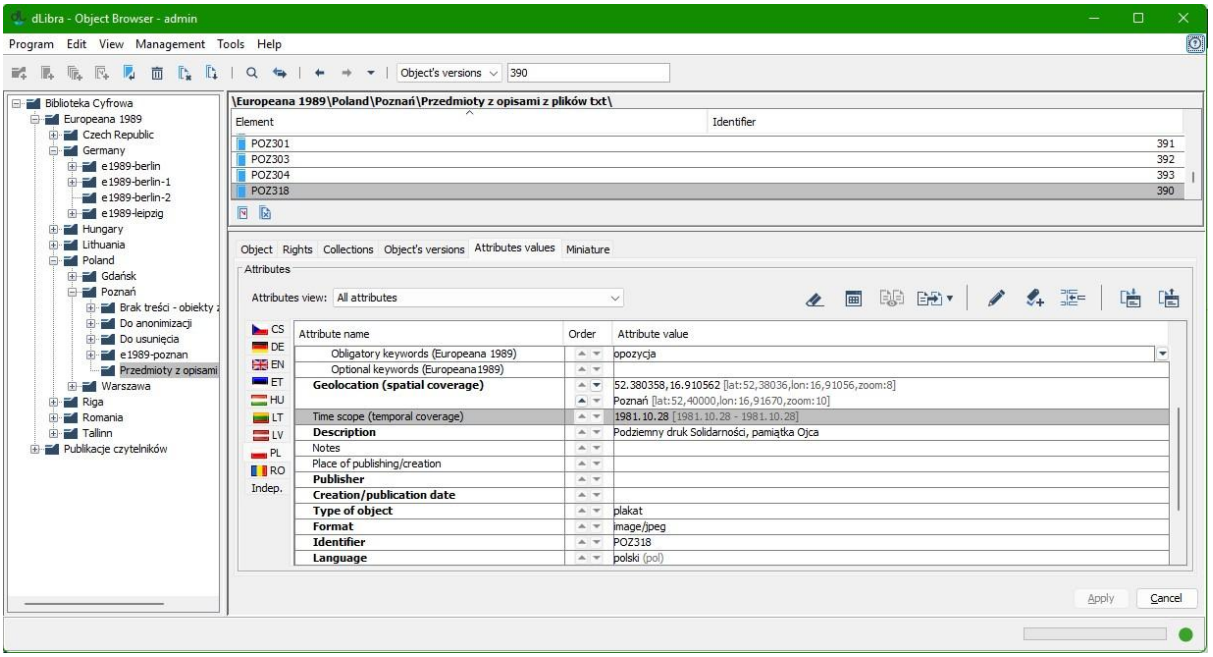

Figure 8. Accepted enrichments integrated with the original record

The REST requests used for transferring the enrichments into the content provider and/or national aggregator software (i.e. Phase 2) are presented in the following subsections.

Fetch enrichments for domain (DEI API)

The following request is used to fetch enrichments available in DEI for records with identifiers matching the given domain. The request is dedicated for dLibra clients exposing their records using certain domain which is part of the OAI-PMH identifier, e.g. *oai:europeana1989.eu:123*. The response will contain only enrichments for records from the calling dLibra domain and in a certain state (which defaults to PENDING). The response includes timespans and places for each returned record.

Type	*GET*
Endpoint	*/api/enrichments/domain*
Authorization	*OAuth 2.0 Access Token*
Query Params	*domain* - domain name that is part of the OAI-PMH identifier

	state - state of the requested enrichments, by default the value is PENDING, this parameter is optional
Response	*200 OK* *400 Bad Request* *404 Not found*

Example

Request

https://fresenia.man.poznan.pl/dei/api/enrichments/domain?domain=europeana1989.eu

Response (shortened)


```
[
        {
        "recordId": "/135/_nnVvT3m",
        "externalId": "oai:europeana1989.eu:1732",
        "timespans": [
        {
                "id": 25193,
                "attribute": "dc:date",
                "item": {
                "itemURL":
"https://europeana.fresenia-dev.man.poznan.pl/dev/documents/story/item?item=434990",
                "pageNo": 1
                },
                "begin": "1988-12-31",
                "end": "1988-12-31"
        },
        {
                "id": 25195,
                "attribute": "dc:date",
                "item": {
                "itemURL":
"https://europeana.fresenia-dev.man.poznan.pl/dev/documents/story/item?item=434991",
```

```
                "pageNo": 2

                },

                "begin": "1988-12-31",

                "end": "1988-12-31"

        },

        {

                "id": 25283,

                "attribute": "dc:date",

                "begin": "1988-12-31",

                "end": "1988-12-31"

        }

        ],

        "places": [

        {

                "id": 25194,

                "attribute": "dcterms:spatial",

                "item": {

                "itemURL":
"https://europeana.fresenia-dev.man.poznan.pl/dev/documents/story/item?item=434990",

                "pageNo": 1

                },

                "latitude": 53.5,

                "longitude": 14.0,

                "name": "Pasewalk",

                "language": "de",

                "wikiDataURL": "https://www.wikidata.org/wiki/Q493924",

                "zoom": 10

        }

        ]

        }

]
```


The JSON structure indicated above contains the enrichments available for multiple records. For each record there is *recordId* with Europeana identifier, *externalId* with OAI-PMH identifier and two separate sections with enrichments for dates (*timespans*) and places (*places*). The *timespans* section contains date enrichments found on certain pages of the record. For such enrichments the item link is provided. There is also one date enrichment without the item link and this contains the timespan with the earliest and latest dates found in the enrichments. The *places* section contains enrichments for locations found on record pages. Each place enrichment

contains geographical coordinates, the name and / or WikiData URL. Language of the name is equal to the language of the document. All types of enrichments also contain the name of the attribute that should be used when applying them to the original metadata.

Fetch enrichments for record (DEI API)

This request is used to fetch enrichments available in DEI for a record identified by the provided *recordId*. By default only enrichments in the PENDING state are returned. The response includes timespans and places for the requested record, using the same structure as indicated in the previous subsection.

Type	*GET*
Endpoint	*/api/enrichments*
Authorization	*OAuth 2.0 Access Token*
Query Params	*recordId* - Europeana record identifier *state* - state of the requested enrichments, by default the value is PENDING, this parameter is optional
Response	*200 OK* *400 Bad Request* *404 Not found*

Example

Request
https://fresenia.man.poznan.pl/dei/api/enrichments?recordId=/615/_8s463v65m
Response (shortened)
{ "recordId": "/615/_8s463v65m",

```
        "externalId": "10.7486/DRI.8s463v65m",
        "timespans": [
        {
        "id": 25286,
        "attribute": "dc:date",
        "item": {
                "itemURL":
"https://europeana.fresenia-dev.man.poznan.pl/dev/documents/story/item?item=39377996",
                "pageNo": 1
        },
        "begin": "1827-12-31",
        "end": "1828-12-30"
        },
        {
        "id": 25288,
        "attribute": "dc:date",
        "begin": "1827-12-31",
        "end": "1828-12-30"
        }
        ],
        "places": [
        {
        "id": 25287,
        "attribute": "dcterms:spatial",
        "item": {
                "itemURL":
"https://europeana.fresenia-dev.man.poznan.pl/dev/documents/story/item?item=39377996",
                "pageNo": 1
        },
        "latitude": 53.3425,
        "longitude": -6.27896,
        "name": "Meath Row, Dublin (since renamed)",
        "language": "ga",
        "zoom": 10
        }
        ]
}
```

Send enrichments validation result (DEI API)

This request is used for informing DEI about enrichment validation results. The validation decision is provided for each record separately but multiple records may be included in the request body. Both accepted and rejected enrichments may be included in the request. This request may be used for records identified by *recordId* which is Europeana identifier or *externalId* which is an OAI-PMH identifier.

Type	*POST*
Endpoint	*/api/enrichments/validate*
Authorization	*OAuth 2.0 Access Token*
Query Params	*JSON body including validation result for each enrichment*
Response	*200 OK* *400 Bad Request* *404 Not found*

Example

Request
https://fresenia.man.poznan.pl/dei/api/enrichments/validate
Request body

[
{
"externalId": "oai:bibliotekacyfrowa.eu:45871",
"timespans": {
"accept": [90, 92, 97],
"reject": [94]
},

```
        "places": {
        "accept": [91,93,96],
        "reject": [95]
        }
  }
]
```

The structure of the JSON provided in the request body must contain the external id of the record (OAI-PMH identifier) and sections *timespans* and *places* with lists of accepted and rejected enrichments identified by the enrichment id. For records identified by the Europeana identifier the field *recordId* may be provided instead of *externalId*.

Development in DRI Repository

The DRI Repository[4] provides long-term preservation of data for the Data Providers, in addition to aggregation to Europeana, and many Irish Data Providers do not have their own platforms. The user interface for reviewing and approving/rejecting enrichments is therefore built into the DRI Platform via a user interface which is made available to the Data Providers. Once enrichments have been reviewed by the data providers, they can export the updated metadata records.

A new menu option to “Fetch Transcribathon Data” was made available to Data Provider users of the DRI platform which creates an asynchronous background job on the server (see Figure 9). This background job fetches all transcriptions and enrichments for the given object. This is done via a call first to the Transcribathon’s Stories API which retrieves a list of all items of a story, followed by a call to the Transcribathon’s Items API which retrieves the available transcriptions and enrichments.

[4] https://repository.dri.ie/

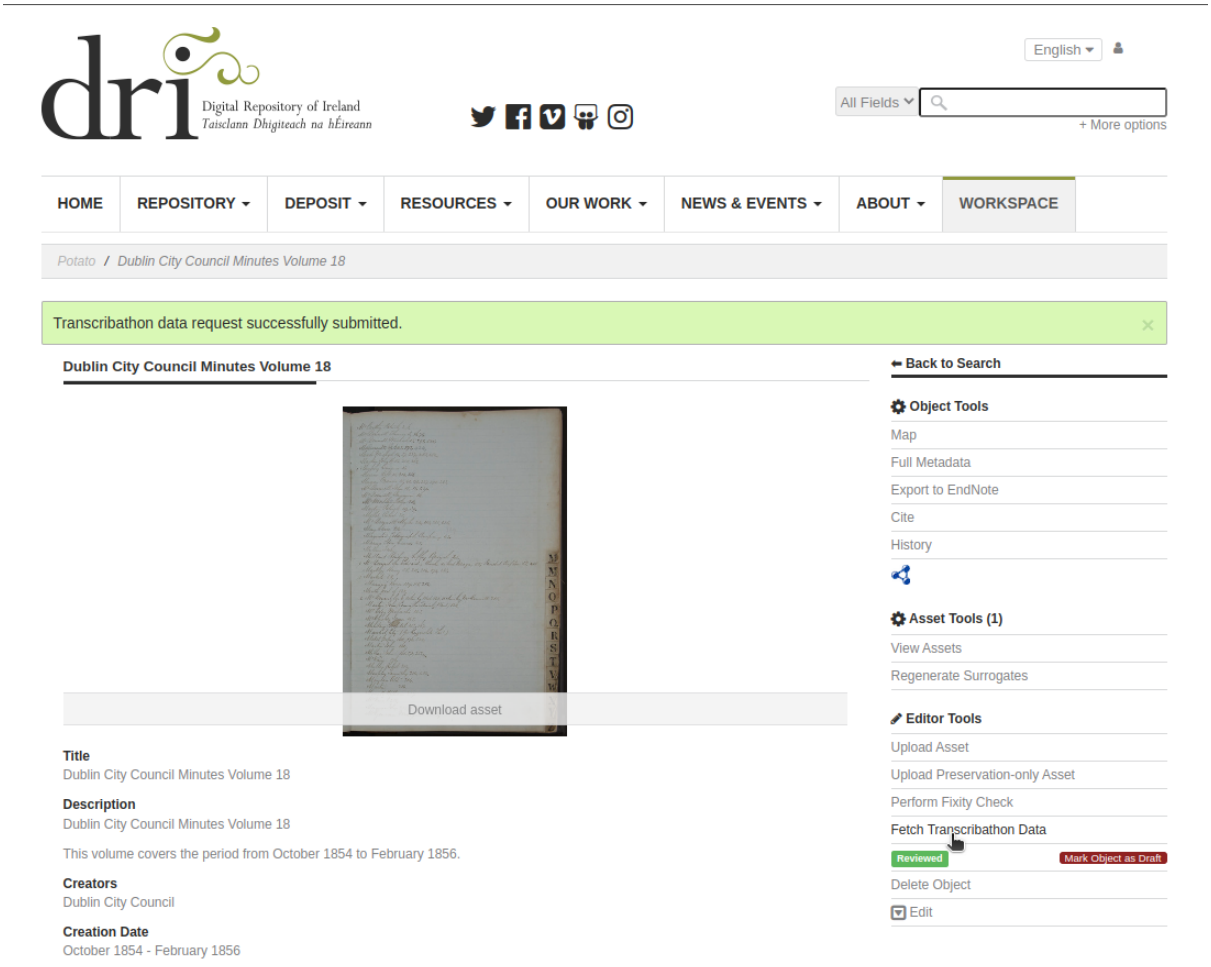

Figure 9: The DRI User Interface to request transcriptions and enrichments from Transcribathon

Enrichments and Transcriptions are initially stored in a custom database until they can be reviewed by the Data Provider. A new “Review Transcribathon Enrichments” page on the DRI platform allows Data Providers to review enrichments and to indicate which of these should be included into the authoritative metadata record.

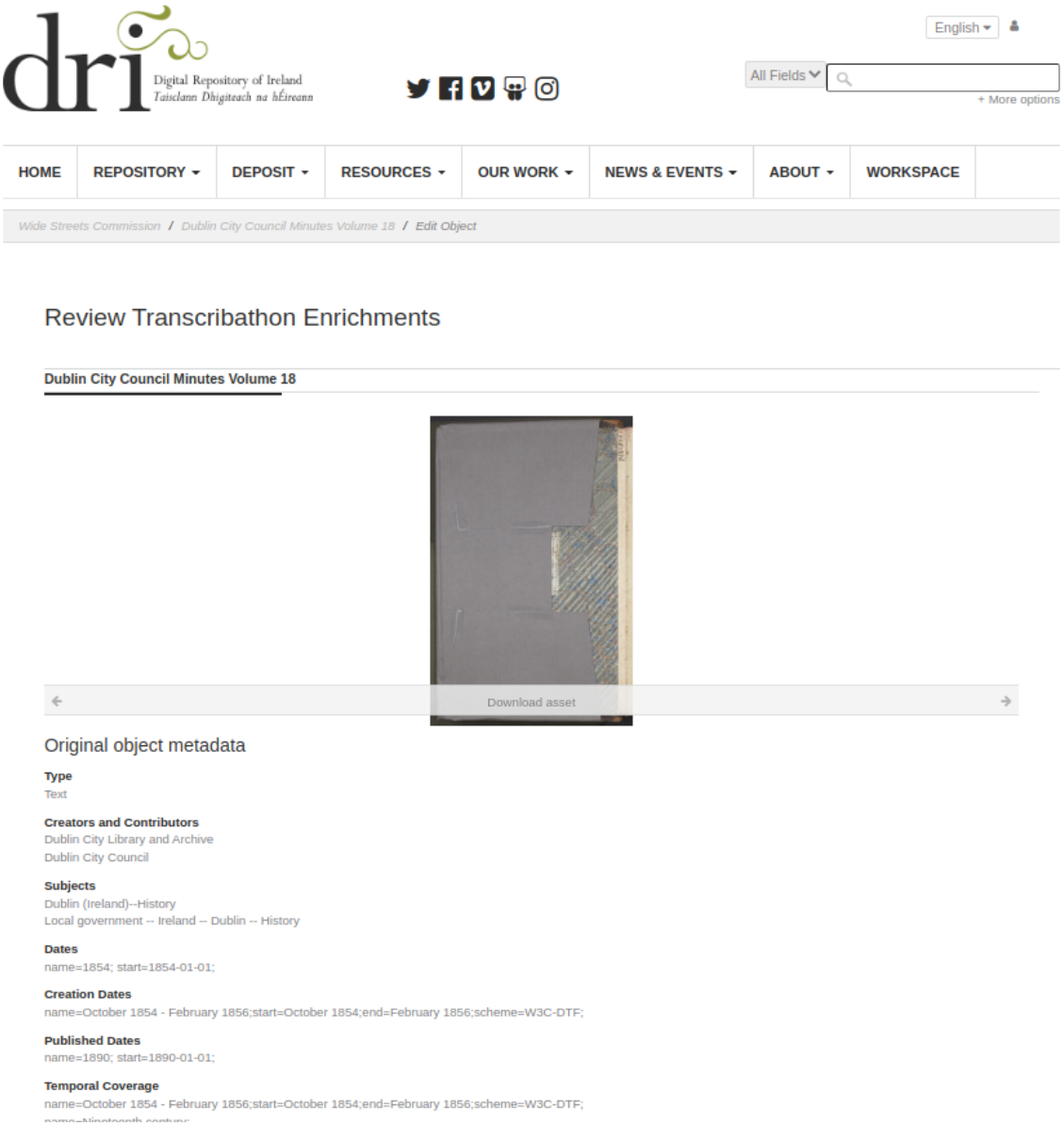

Figure 10: The DRI "Review Transcribathon Enrichments" page

When initially landing on the Review Transcribathon Enrichments page, the Data Provider will be able to view the digital object and the main metadata fields (see Figure 10). This allows them to compare the existing metadata with the enrichments.

On scrolling down the Review Transcribathon Enrichments page, the Data Provider is presented with the individual enrichments, and for each one, has the option to accept. For each enrichment, a link to the IIIF image is provided to allow the Data Provider to verify the enrichment directly if required (see Figure 11).

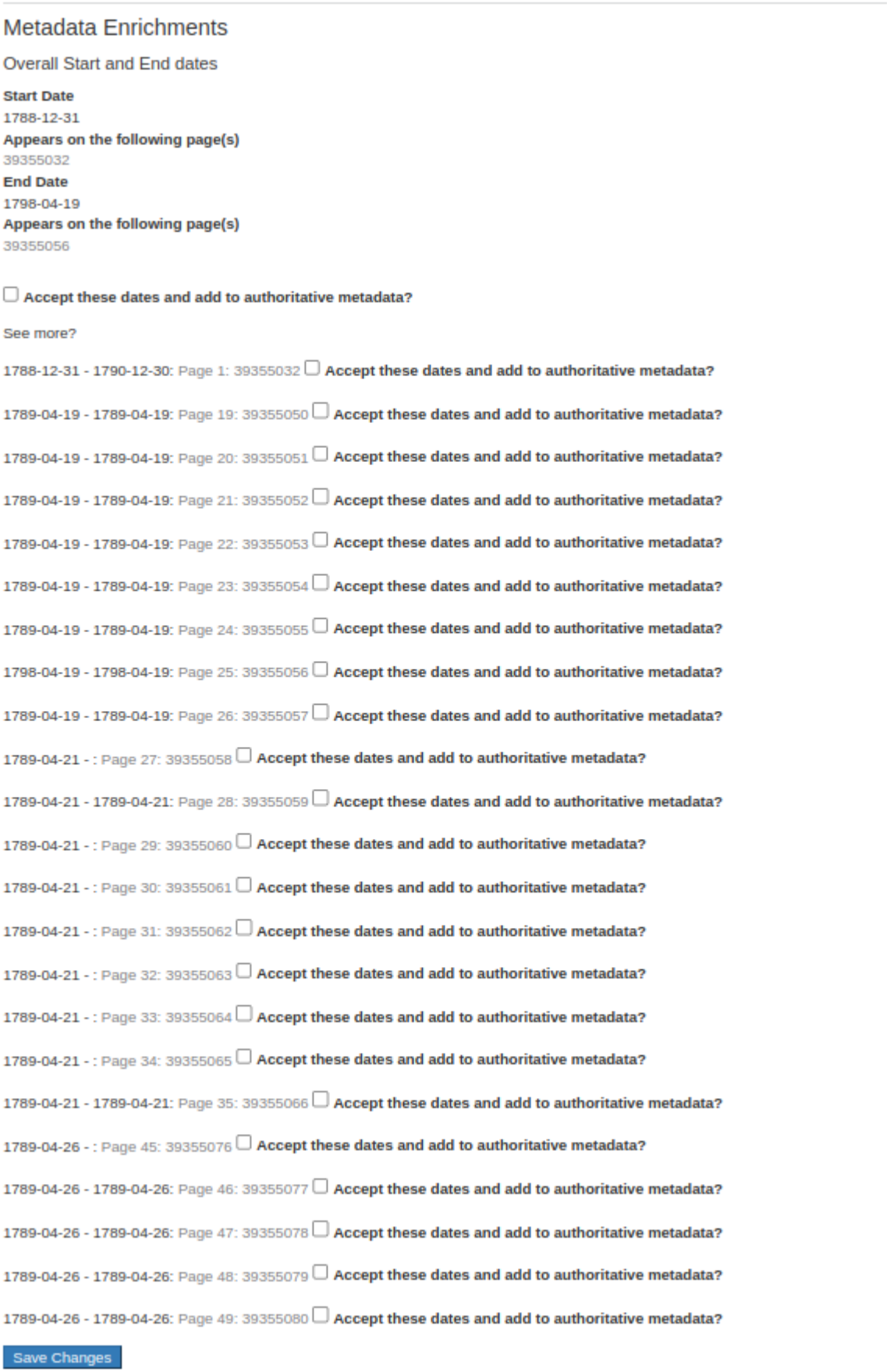

Figure 11: The DRI interface to accept enrichments and save them to the authoritative metadata

Phase 3 - reaggregation of enriched records into Europeana

This phase also differs slightly in case of dLibra instances and DRI national aggregator. Once the enrichments are accepted in dLibra the updated records have to be reaggregated by the FBC national aggregator. This is the place where conversion from dLibra metadata schema to EDM schema is applied. During this operation the newly added values are included in the EDM record. In some cases they result in the creation of contextual classes which might improve the metadata tier of the record in Europeana. The conversion process is already available in the aggregation software and will not be described in detail. The last step of this phase is the (re-)ingestion to Europeana, which in case of FBC is done by providing ZIP archive with XML files containing separate EDM records. The diagram below illustrates the process of harvesting records from dLibra, applying enrichments to EDM records and their re-ingestion to Europeana.

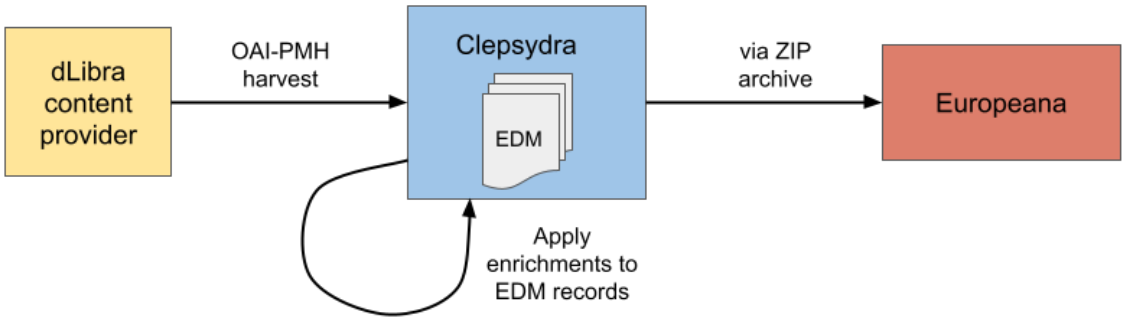

Figure 12. Applying enrichments to records and reingestion to Europeana.

The re-ingestion routine is similar in the case of Irish National Aggregation (i.e DRI). Newly added values are automatically included when the EDM record is exported, including any new contextual classes which might improve the metadata tier of the record in Europeana. This occurs using the already existing aggregation workflow whereby DRI staff use an export tool to create a ZIP archive of the EDM XML files. An OAI-PMH feed of the EDM is also available in the DRI platform. The diagram below shows this workflow.

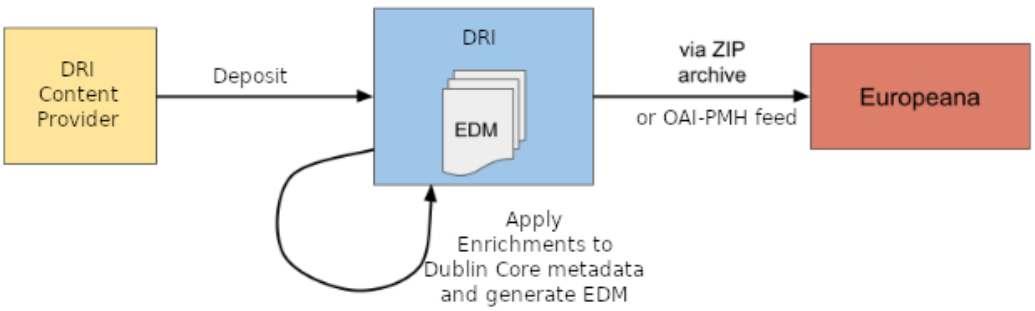

Figure 13. Applying enrichments and reingestion to Europeana in DRI

Outlook and future improvements

Current implementation of the UI in the Transcribathon Platform does not allow users indicating the meaning of the dates they add to the records. Therefore, all the date enrichments are related to the general *dc:date* attribute. The next step of improving enrichment possibilities is to use more specific dates like creation date or issue date. That would mean adding dates to *dcterms:created* and *dcterms:issued* attributes. Although adding more *edm:TimeSpan* contextual classes might not increase metadata tier in Europeana such enrichments will help Europeana users searching for materials in a more precise way, enabling more advanced search scenarios in Europeana.

Another very important feature is handling the information about persons mentioned in the transcribed materials. Transcribathon currently allows contributing names, place and date of birth and death of the persons associated with the record. Further work in this area will focus on indicating the role of the person in context of the record and exposing that information in the Transcribathon Platform API. The contributed enrichments will be applied to *dc:creator*, *dc:contributor* or *dc:subject* attributes. This will allow creating *edm:Agent* contextual classes which have great value for Europeana users.

All above features will be covered in the Transcribathon Platform API. This will also imply changes in Data Exchange Infrastructure, dLibra and DRI to support new types of enrichments.

Additional work will also be required in order to integrate transcriptions into the National Aggregator and Content Provider platforms.

Conclusions

This document presents the work carried out for achieving the objectives of milestone 7 of EnrichEuropeana+ Action. It describes the current state of work as well as planned improvements that will be developed before the end of the project. The description contains the enrichment validation workflow and the specifications for the API endpoints for all services involved in the enrichment process. The document shows how the crowdsourced enrichments from the Transcribathon tool are delivered to the dLibra and DRI infrastructures for validation and integration into the authoritative metadata. The required enhancements on the side of dLibra and DRI applications are presented. The outcome of this operation will be visible in Europeana after reingestion of the materials involved in this process. This will be done before the end of the project when all necessary features are implemented in Transcribathon, DEI, dLibra and DRI to avoid multiple reingestion of the same materials in Europeana.