



ECP-2008-DILI-528001

EuropeanaConnect

D5.7.1 – EOD Connector – Documentation and Final Prototype

Deliverable number/name	<i>D 5.7.1</i>
Dissemination level	<i>PP (Restricted to other programme participants)</i>
Delivery date	<i>01/05/11</i>
Status	<i>Final</i>
Author(s)	<i>Silvia Gstrein (UIBK), Andreas Parschalk (UIBK), Joachim Jung (AIT)</i>



eContentplus

This project is funded under the eContentplus programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.



D5.7.1 – EOD Connector – Documentation and Final Prototype

Documentation and Design



co-funded by the European Union

The project is co-funded by the European Union, through the **eContentplus** programme

<http://ec.europa.eu/econtentplus>



EuropeanaConnect is coordinated by the Austrian National Library



Table of contents

Abstract	4
1. Introduction	5
1.1 General background: EOD service and network	5
1.2 General background: EOD within WP 5	7
2. Delivery sources of metadata	8
2.1. Libraries with OPAC catalogues.....	8
2.2. Libraries with image catalogues	10
2.3. ODM operator generated orders	12
3. Storage sources of EOD eBooks	13
3.1 Case 1: Repository, OAI-PMH, ESE format, EOD set or no EOD set.....	13
3.2 Case 2: Repository, OAI-PMH, EOD set or no EOD set, but no ESE format.....	14
3.3 Case 3: Website with viewer without harvestable metadata	14
3.4 Case 4: No accessible storage at all	15
4. EOD connector specification	15
4.1 Harvesting and OAI-PMH infrastructure	15
4.1.1 Harvesting a partner with records exposed as ESE via OAI-PMH.....	16
4.1.2 Harvesting a partner with records exposed in other format via OAI-PMH.....	17
4.1.3 Dealing with partners without (harvestable) repository	17
4.1.4 Dealing with partners that are already aggregated by other provider(s) or are Europeana Aggregators themselves	19
4.2 The EOD Repository.....	20

Abstract

The eContentplus project **EuropeanaConnect** delivers core components for Europeana to make available via innovative access channels a vast amount of digital content from libraries, museums, archives, and audio-visual collections from across Europe. In addition EuropeanaConnect integrates key added-value services into Europeana which will significantly enhance its usability and functionality.

One component for the creation of an operational Europeana service which is delivered by EuropeanaConnect is the **e-Books-on-Demand service**, which is the outcome of the eTEN co-funded project “Digitisation on Demand”¹ (2006-2008), where the European market was validated for such a service and the pilot service was extended to the participating libraries. The main objective in task 5.7 within the EuropeanaConnect project is and was to enhance and adapt this EOD service allowing for an automated and generic transfer of the eBooks created within EOD to Europeana.

The objective of this paper is to document and describe the design of a prototype connection mechanism between the e-Books-on-Demand service (EOD) and Europeana, which was implemented from M7 to M24 under the working name: “EOD connector”. Since EOD is currently dealing with 30 different libraries from 12 European countries (<http://books2ebooks.eu/partner.php5>) it has the role of an aggregator for Europeana for EOD eBooks. The main approach is that the interface must allow for a flexible aggregation from very different sources, such as, in the best case, OAI-PMH repositories, but also from libraries with e.g. conventional homepages containing book viewers for each digital object or no storage and display of eBooks at all.

In the first part, the paper will give an overview of the EOD service and the libraries involved. In the following parts, the core issues for the aggregation of EOD eBooks will be described, such as the different metadata formats and repositories. In the last part, the design of the final prototype of the “EOD connector” will be documented.

1

http://ec.europa.eu/information_society/activities/eten/cf/opdb/cf/project/index.cfm?mode=detail&project_ref=ETEN-518635

1. Introduction

1.1 General background: EOD service and network

The eBooks on Demand (EOD)² network enables users to order digital copies of copyright-free books from the EOD network libraries' collections. The libraries digitise the requested items and deliver them to the end users via the EOD service network. The end user receives the eBook in the form of a PDF with the OCR full text in the background (if applicable). The libraries store the original masterfiles, the generated PDF files and (if applicable) the Abbyy output XML and RTF files. The books digitised in this way are incorporated into the digital libraries and repositories of the participating libraries after a certain retention period (about 2-3 months) and are thus also accessible on the Internet.

eBooks on Demand from a readers' perspective:



All books available for EOD are marked in the library catalogues with this label. Whenever a user finds a catalogue entry with this sign, s/he is able to order this book as an EOD eBook and it will be digitised on his/her request.

The libraries offering EOD may integrate the EOD button in as many library catalogues as they wish. Usually, a library tries to implement the service in the local library catalogue and a union or shared catalogue such as national or regional union catalogues (e.g. Austrian union catalogue or Bavarian Library Network) or pan-European catalogues such as The European Library catalogue (see Fig. 1.).

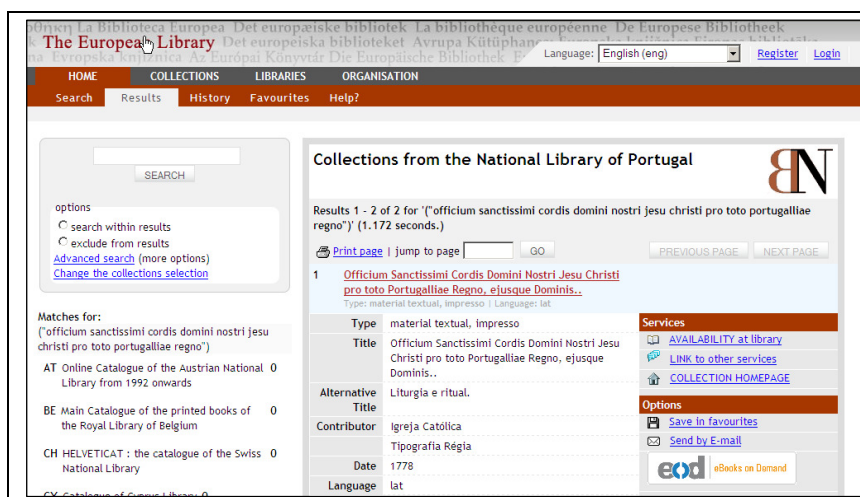


Fig.1: Example of book record from the National Library of Portugal with EOD button in the TEL catalogue

² <http://books2ebooks.eu>

eBooks on Demand from a libraries' perspective:

The EOD service is implemented within the framework of a network. The individual library is responsible for the processing of the order, the digitisation of the book and the long term preservation. The creation of eBooks, delivery to the customer, electronic payment and automatic text recognition (OCR) even for gothic font texts are supported through a central database with web access.

Participating libraries:

Currently, over 30 libraries from 12 European countries participate in the EOD network and offer the service. Starting originally from 5 national libraries and 8 university libraries, meanwhile also research libraries, academy of sciences libraries and public libraries take part. The network aims at enlarging the service in order to be able to provide digitisation on demand in as many countries and libraries as possible. The goal is to double the number of network members within the next 4 years. It is expected that other 15 to 20 libraries will join the network during the next 2 to 3 years.

Current members offering EOD:

<p>Austria</p> <ul style="list-style-type: none"> • University of Innsbruck, Library (co-ordinator) • University Libraries of Graz, Salzburg and Vienna, Library of the Medical University of Vienna • Vienna City Library • St. Pölten Diocese Archive <p>Germany</p> <ul style="list-style-type: none"> • University Libraries of Regensburg, Greifswald, Leipzig and Humboldt-Universität zu Berlin • Bavarian State Library • Saxon State and University Library (Dresden) <p>Denmark</p> <ul style="list-style-type: none"> • The Royal Library <p>Estonia</p> <ul style="list-style-type: none"> • National Library • University Library of Tartu <p>France</p> <ul style="list-style-type: none"> • Academic Health Library (Paris) <p>Sweden</p> <ul style="list-style-type: none"> • Umeå University Library 	<p>Czech Republic</p> <ul style="list-style-type: none"> • Moravian Library in Brno • Research Library in Olomouc • Library of the Academy of Sciences in Prague • National Technical Library <p>Hungary</p> <ul style="list-style-type: none"> • National Széchényi Library of Hungary • Library of the Hungarian Academy of Sciences <p>Portugal</p> <ul style="list-style-type: none"> • National Library <p>Slovakia</p> <ul style="list-style-type: none"> • University Library in Bratislava • Slovak Academy of Sciences <p>Slovenia</p> <ul style="list-style-type: none"> • National and University Library <p>Switzerland</p> <ul style="list-style-type: none"> • The Swiss National Library • The Library at Guisanplatz
---	--

Fig.2: Table of current EOD libraries (April 2011)

1.2 General background: EOD within WP 5

Within EuropeanaConnect WP 5, UIBK is task leader of **Task 5.7**. The main objective in task 5.7 is to enhance and adapt the EOD service to make eBooks created within EOD accessible to Europeana according to the Europeana specifications. Task 5.7 heavily relies on Task 5.3 “Development of the Europeana OAI-PMH Management Infrastructure” for the aggregation of metadata and submission to Europeana. More information on the Europeana OAI – PMH Infrastructure can be found online.³ The Europeana website provides more information on how to contribute metadata and which specifications must be met.⁴

³ http://www.europeanaconnect.eu/documents/01_Europeana_OAI_PMH_Infrastructure.pdf

⁴ <http://version1.europeana.eu/web/europeana-project/provide-content>

2. Delivery sources of metadata

The digitisation of books can be ordered from EOD partner libraries through up to four different channels: OPAC catalogues, digitised card catalogues, free web forms or by contacting the library directly. These different channels result in input of metadata sets of varying quality into the networks' digitisation workflow management software (ODM). The metadata of EOD books are kept as a subset of Dublin Core simple. They are only corrected and validated to the point that is necessary for the eBook generation and delivery process as such, as well as for the generation of the eBook cover pages. This is due to the fact that the additional workload of the ODM operators should be kept to a minimum.

2.1. Libraries with OPAC catalogues

Almost all member libraries of the EOD network have placed the EOD button in their own OPAC or in a union catalogue. In this case the eBooks on Demand customer finds the book to be digitised in the respective catalogue and is presented with the EOD button in the record details. Clicking this button will forward the customer to the EOD order form, which is already prefilled with the basic metadata of the ordered book. In this case the metadata of the book are fetched via z39.50 and converted from their original format (MAB, MARC,...) to the ODM internal metadata format. The customer is not able to edit the book data. The following screenshot shows the EOD button at a sample record from the OPAC⁵ at the Saxon State and University Library (SLUB) in Dresden.

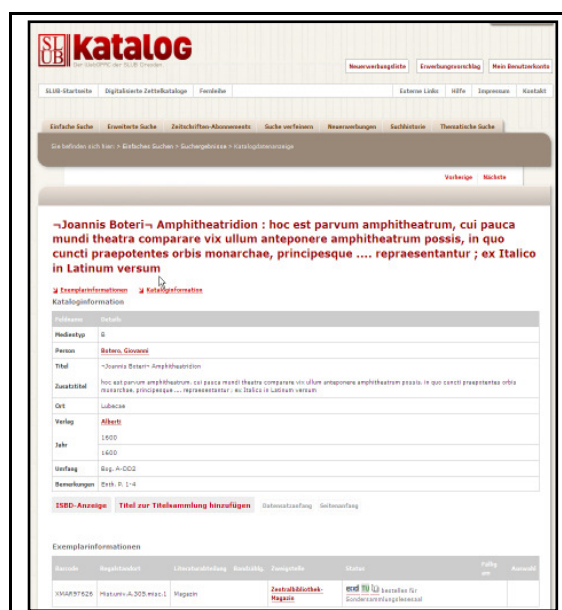


Fig.3: SLUB OPAC record with EOD button

This order process results in reliable metadata fetched from the libraries' (or union) catalogue directly as seen in the following screenshot:

5 <http://webopac.slub-dresden.de/>

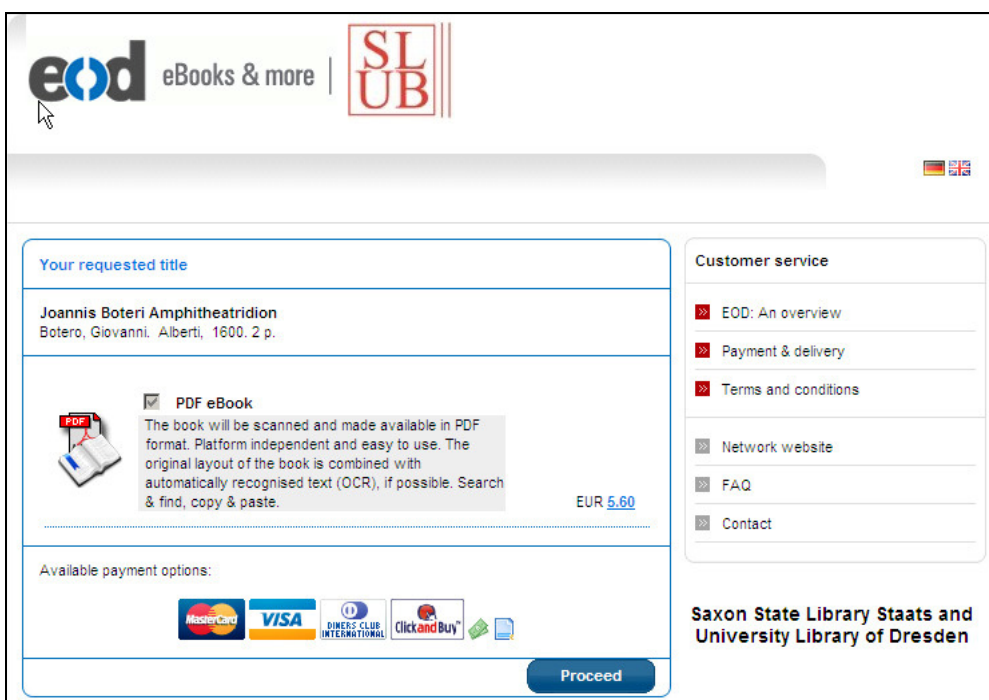


Fig.4: Order form with book metadata fetched via Z39.50

Then the metadata is matched to Dublin Core Simple and may in addition be altered later in the digitisation workflow process by the ODM operator:

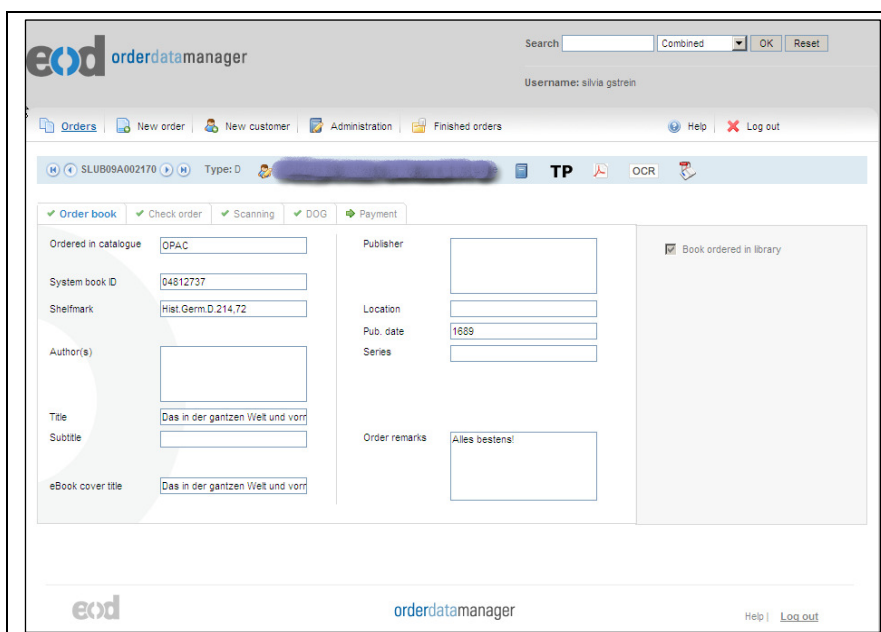


Fig.5: Snapshot from the Order Data Manager (ODM), step “order book” with book metadata

2.2. Libraries with image catalogues

Many of the EOD partner libraries have a digitised card catalogue, where they place the EOD button with every single card image.

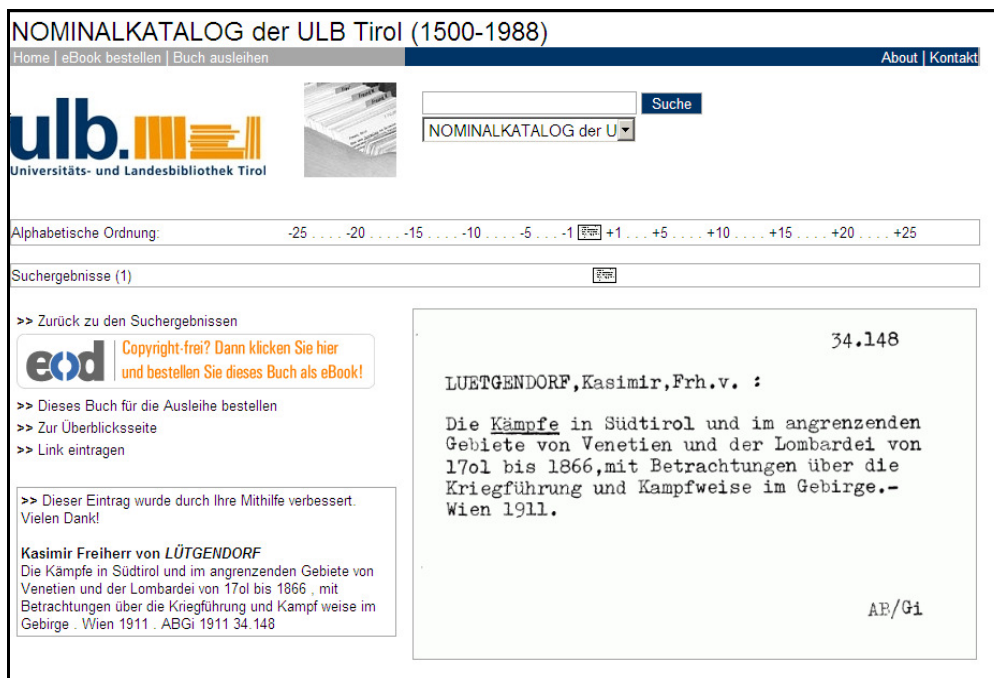


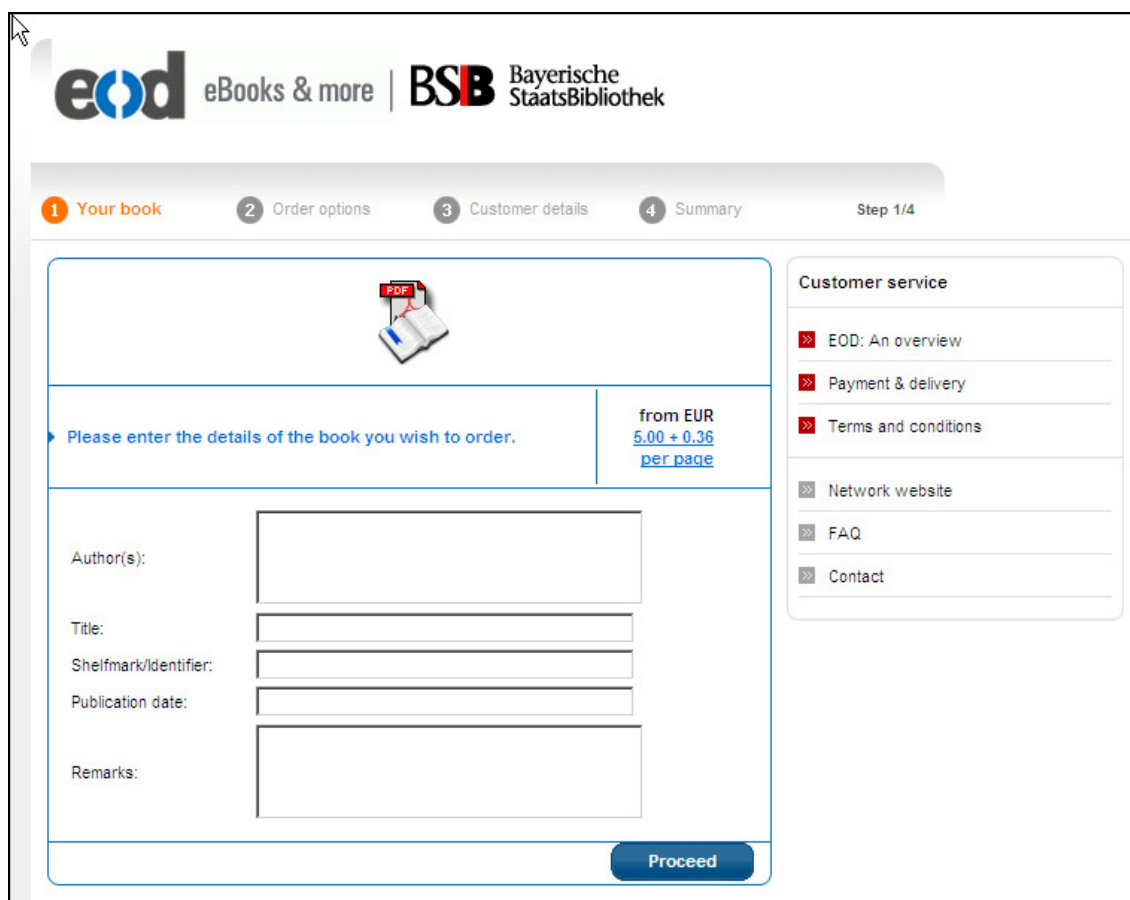
Fig.6: Digitised card catalogue from Innsbruck University library with a sample record and the EOD button

After clicking this EOD button the customer is forwarded to the EOD order form and is shown an empty form and the card from the catalogue. The customer is asked to carry over the record data and proceed with the order:



Fig.7: EOD order form with digitised card as metadata

Additionally each library has the possibility to use a “general web form”. This is a simple variant of the EOD order form, where the user can request digitisation of a book without having found it in a catalogue first:



The screenshot shows a web form for ordering eBooks. At the top, the logo for 'eod eBooks & more' and 'BSB Bayerische Staatsbibliothek' is visible. Below the logo is a progress bar with four steps: 1. Your book (active), 2. Order options, 3. Customer details, and 4. Summary. The current step is 'Your book', which contains a PDF icon and a prompt: 'Please enter the details of the book you wish to order.' To the right of this prompt, the price is listed as 'from EUR 5.00 + 0.36 per page'. The form fields include: Author(s), Title, Shelfmark/Identifier, Publication date, and Remarks. A 'Proceed' button is located at the bottom right of the form. On the right side of the page, there is a 'Customer service' sidebar with links for: EOD: An overview, Payment & delivery, Terms and conditions, Network website, FAQ, and Contact.

Fig.8: Example of a “general web form” for free metadata entry

Both of these metadata input methods do not result in fully reliable data as the only mandatory field is the title field and in most cases have to be corrected and enhanced by the ODM operator to fulfil the minimum standard required for the eBook generation process.

2.3. ODM operator generated orders

ODM operators can directly create orders from within the ODM. This feature is used frequently for in-house orders or digitisation requests via telephone or e-mail. The metadata is either entered manually by the operator or from v2.0 of the ODM on (release 11/2009) can be fetched via z39.50 from the library's OPAC catalogue if a system ID is known.

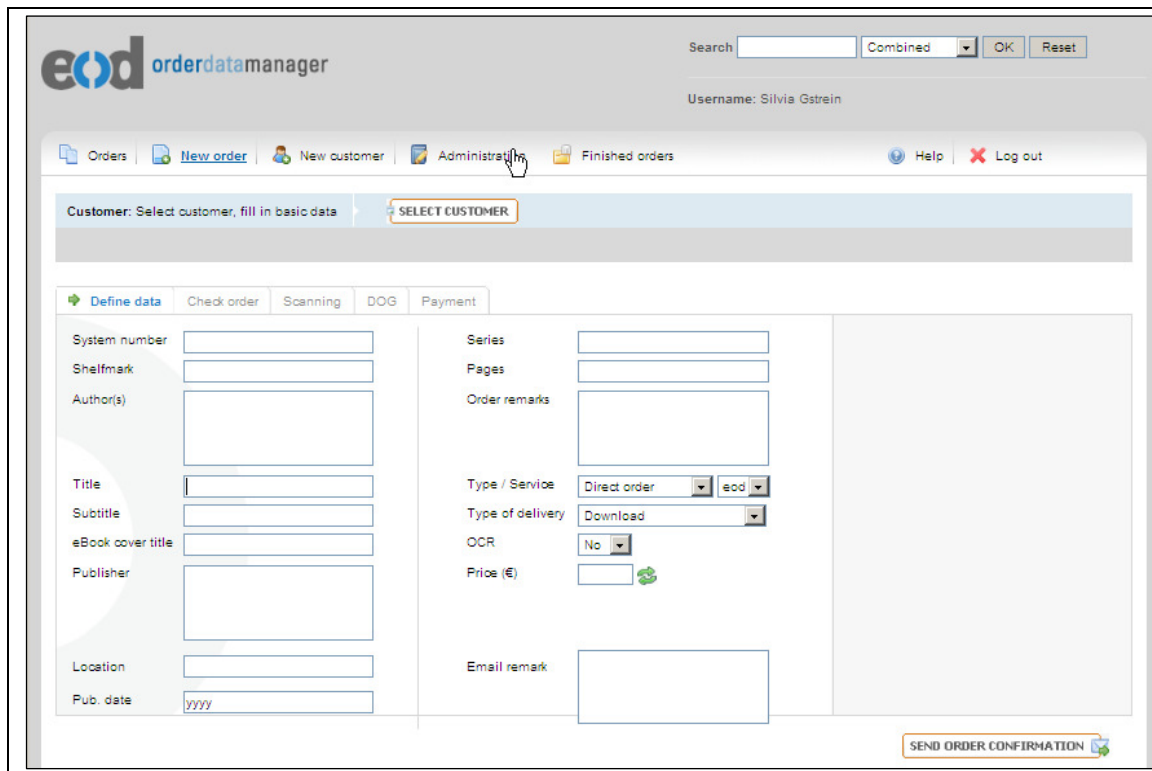


Fig.9: “New order” within the Order Data Manager (ODM) for creation of orders directly by ODM operators

Metadata entered via this channel are generally more complete and correct than data entered by customers as ODM operators are mostly librarians and are trained in metadata matters, but will still usually only comply with the minimum requirements for the eBook generation process.

3. Storage sources of EOD eBooks

After having processed the orders via the ODM and after successful payment, the customer downloads the PDF file. Then, after 2 – 3 months (the so called “retention period”), the EOD libraries integrate the digital object into their local repositories (if available). The “EOD connector” built in task 5.7.2 harvests metadata from those repositories in order to make available EOD eBooks to Europeana. This chapter explores the different conditions for making harvesting possible.

3.1 Case 1: Repository, OAI-PMH, ESE format, EOD set or no EOD set

Three EOD member libraries are currently able to make the metadata to their EOD eBooks available to UIBK via OAI-PMH in ESE format. These are the University Library of Regensburg, National and University Library in Slovenia and the Moravian Library (from 2011 on).



The screenshot shows the Dlib.si website interface. At the top left is the logo 'Dlib.si Digitalna knjižnica Slovenije'. At the top right are navigation links: 'collections | portal | news | contacts | help | slovene'. The main heading is 'Results for collection: Books'. Below this is a search filter section with 'Sort by:' and options for 'author', 'title', and 'year'. A list of seven search results is displayed, each with a title, year, file format and size, and a note that the material is digitalized within the EOD service. On the right side, there are search statistics: 'Search results: 70 Books', 'Search history: EOD', and options to 'Refine search' or 'New search'.

Rank	Title	Year	File Format & Size	Material Contributed by
1.	Spomenica Miroslavu Vilharju	1906	pdf document (2134 kB) / html preview (63 kB)	Digitalizirano v okviru storitve EOD
2.	Burke in porednosti (sramežljive kratkočasnice)	1914	pdf document (4005 kB) / html preview (217 kB)	Digitalizirano v okviru storitve EOD
3.	Sanje v podobah po najboljših virih zbrane	1902	pdf document (4101 kB) / html preview (188 kB)	Digitalizirano v okviru storitve EOD
4.	Prave egiptovske sanjske bukve od leta 1231. z razkladanjem sanj in s pristavljenimi numerami za ...	1901	pdf document (1803 kB) / html preview (68 kB)	Digitalizirano v okviru storitve EOD
5.	Greh in smeh ... (zabeljene in osoljene kratkočasnice)	1912	pdf document (3173 kB) / html preview (237 kB)	Digitalizirano v okviru storitve EOD
6.	Marja Pomagaj na Brezjah in čudne dogodbe na tem kraju	1887	pdf document (1209 kB) / html preview (26 kB)	Digitalizirano v okviru storitve EOD
7.	To je tapravi inu tazieli Colemone-Shegen (kateri je biv vkeleranje naprvevo bart vdrvkan vtam le...	1800	pdf document (20 MB) / html preview (267 kB)	Digitalizirano v okviru storitve EOD

Fig.10: Screenshot of the repository from the National and University Library of Slovenia with the section of EOD eBooks (<http://www.dlib.si/>)

3.2 Case 2: Repository, OAI-PMH, EOD set or no EOD set, but no ESE format

Approximately another third of the current member libraries dispose of a repository which is harvestable via OAI-PMH, but the metadata won't be made available in ESE format, but rather in Dublin Core, DC TEL Application profile or MARC21. Libraries which fall under this category are mainly national libraries such as the National Library of Estonia.



Fig.11: Screenshot of the digital repository from the National Library of Estonia with a sample of an eBook which was digitised via EOD (<http://digar.nlib.ee/otsing/avaleht>)

3.3 Case 3: Website with viewer without harvestable metadata

Some libraries present their EOD eBooks on especially dedicated websites where in most cases metadata is presented together with the link to the download of the PDF or a book viewer.

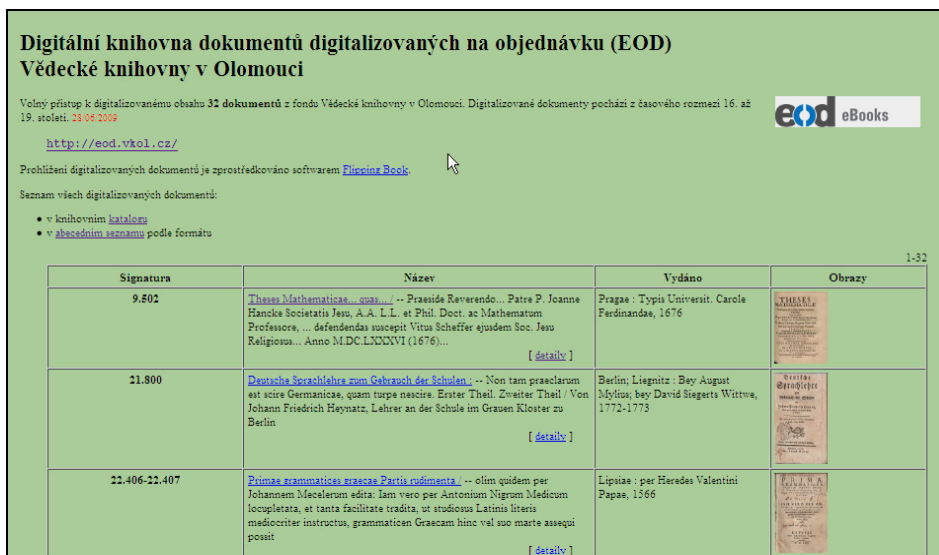


Fig.12: Screenshot of website containing metadata of books digitised via EOD together with links to the “flipping book” viewer (<http://eod.vkol.cz/>)

3.4 Case 4: No accessible storage at all

Some libraries don't have currently the possibility to make their EOD eBooks publicly available at all and won't be able in short term either as this is the case for e.g. the Vienna City Library. Thus, for those libraries and most probably also for the libraries from case 3, an "interim" solution needed to be found. However, on the long term, also these libraries will most probably host their own repository or contribute their digital objects to some kind of shared repository.

From the considerations of this chapter, the following conclusions were drawn concerning the validation of metadata:

- Cases 1+2: Metadata may be retrieved via OAI-PMH from repository
- Cases 3+4: Metadata needs to be controlled in the ODM by operators

4. EOD connector specification

4.1 Harvesting and OAI-PMH infrastructure

UIBK aggregates metadata from EOD eBooks from repositories of EOD member libraries and exposes it to Europeana. Therefore, UIBK deployed an instance of REPOX⁶ to aggregate all harvestable OAI-PMH interfaces of the EOD networks' partner libraries. REPOX is a framework development in WP5 of EuropeanaConnect to manage metadata spaces and is used to harvest the metadata, convert from differing formats of metadata when needed as well as expose the sum of the harvested records via OAI-PMH to Europeana. REPOX supports conversion from various formats to oai_dc and tel application profile and is expandable to support other conversions with XSLT. Metadata Crosswalks have been created to be able to expose the harvested records to Europeana via OAI-PMH in the current version of ESE.

Each partner library harvested by UIBK has its own set. The sets are named according to the following scheme: EuropeanaConnect_<library acronym>, e.g "EuropeanaConnect_UBER" in the case of the library of Humboldt Universität zu Berlin. The metadataPrefix for OAI-PMH harvesting is called "ESE".

Already submitted and ingested metadata can be found using the following URL:

http://www.europeana.eu/portal/search.html?query=europeana_collectionName%3A08902*

Taking into account the considerations from the previous chapter, we can discern four basic cases for harvesting the necessary metadata:

- The partner library is able to expose the repository metadata in ESE via OAI-PMH
- The partner library is able to expose the repository metadata in a different format via OAI-PMH
- The partner library does not expose the metadata of digitised eBooks via an easily harvestable interface
- The partner library is already aggregated by other provider(s) or is Europeana Aggregator itself

⁶ <http://repor.ist.utl.pt/>

The premise for the first two cases is that the partner library has defined a set “eod” within the oai-pmh harvestable repository. Using the set “eod” for EOD eBooks only and a different set for other collections will allow avoiding duplicates in Europeana.

In the following sub chapters, the different scenarios for harvesting metadata will be described in detail.

4.1.1 Harvesting a partner with records exposed as ESE via OAI-PMH

This is the trivial case and affects currently 3 libraries in the EOD network. The partners' repository is harvested, and the <europeana:provider> and <europeana:dataProvider> elements are filled using XSL transformations of the harvested metadata within REPOX. The <europeana:provider> field is set to “EuropeanaConnect – EOD”, the <europeana:dataProvider> field is filled with the harvested library's name. In case the library has already submitted other records to Europeana the institution's name as currently shown in the Europeana portal will be used.

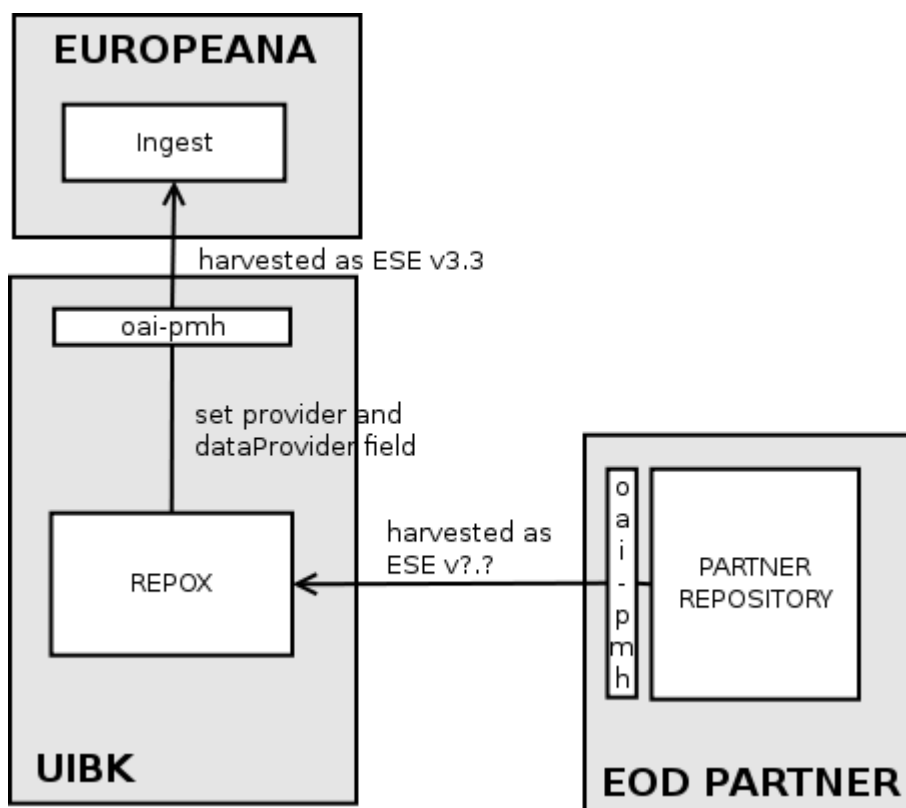


Fig. 13: Harvesting repositories with OAI-PMH (ESE)

4.1.2 Harvesting a partner with records exposed in other format via OAI-PMH

This case can also be implemented with the functionality offered by REPOX. The metadata is harvested in the existing format and is converted to ESE using XSLT. Metadata crosswalks have been implemented for oai_dc, marc21, DC TEL application profile and earlier versions of ESE to ESE v3.3.

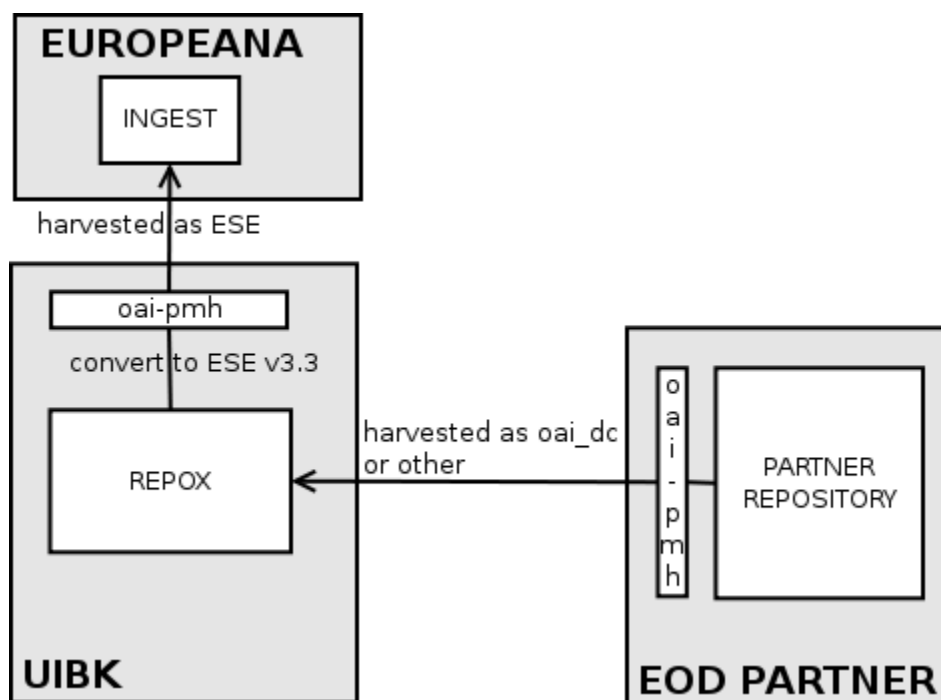


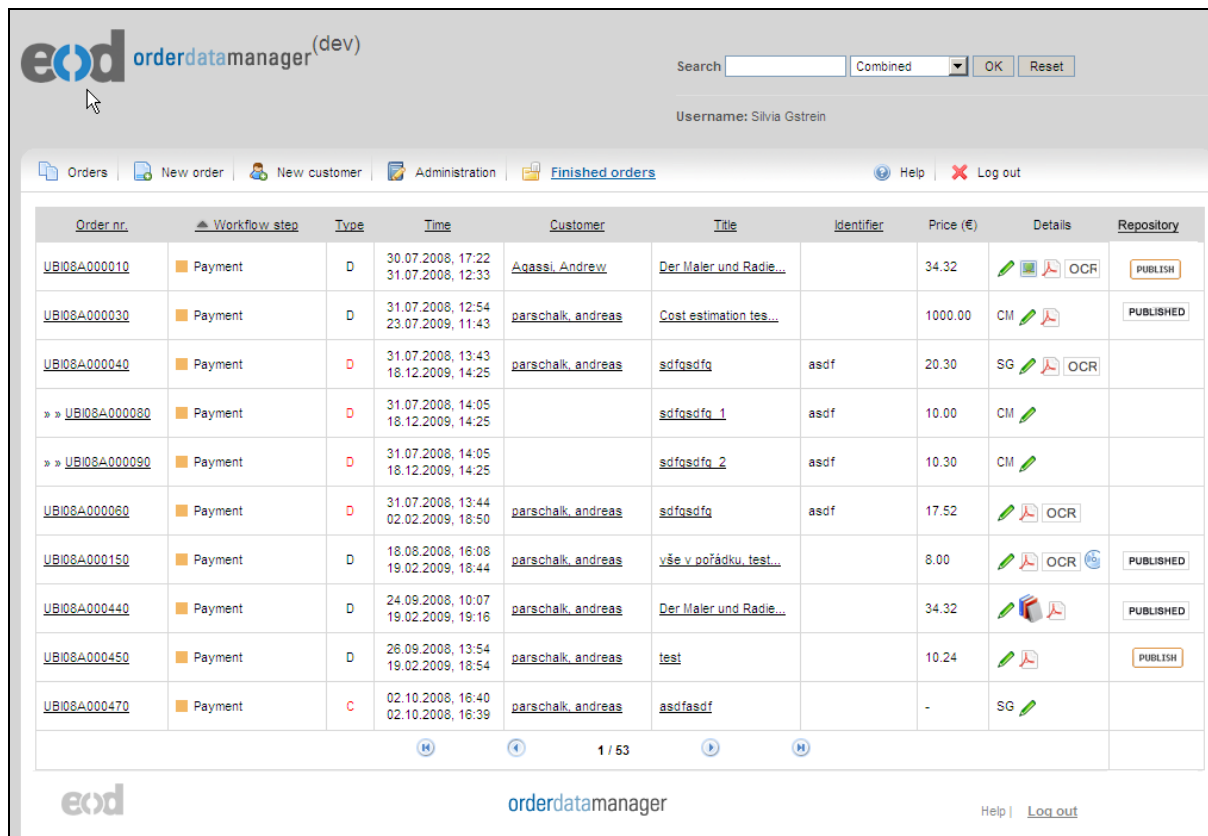
Fig. 14: Harvesting repositories with OAI-PMH (other formats)

4.1.3 Dealing with partners without (harvestable) repository

Partners of the EOD network, which do not expose their metadata through OAI-PMH, or do not offer a harvestable repository at all, are offered a possibility to make available their digitised content to Europeana within the EOD workflow software Order Data Manager (ODM).

Once a partner decides to make available an item previously processed within the ODM, the metadata held within the software first has to be validated manually before being submitted to Europeana. This is due to the fact that correctness of the metadata held within the ODM cannot be guaranteed (cf. 2c metadata input within ODM).

In the "finished order" overview of the ODM (see Figure 15) the operator will be informed about the possibility to make available an eBook to Europeana after the retention period (cf. 1.1) has expired. Once this is accepted, the operator is shown a form, where the metadata of this eBook (as dublin core) is prefilled with the values available in the ODM. If this record was ordered from within an OPAC and a z39.50 interface is available, the record can also be re-fetched from the catalogue.



The screenshot shows the 'eod orderdatamanager' web application. At the top, there is a search bar and a dropdown menu set to 'Combined'. Below that, the user is identified as 'Silvia Gstrein'. The main navigation bar includes 'Orders', 'New order', 'New customer', 'Administration', and 'Finished orders' (which is highlighted). There are also 'Help' and 'Log out' buttons.

Order nr.	Workflow step	Type	Time	Customer	Title	Identifier	Price (€)	Details	Repository
UBI08A000010	Payment	D	30.07.2008, 17:22 31.07.2008, 12:33	Agassi, Andrew	Der Maler und Radie...		34.32		PUBLISH
UBI08A000030	Payment	D	31.07.2008, 12:54 23.07.2009, 11:43	parschalk, andreas	Cost estimation tes...		1000.00	CM	PUBLISHED
UBI08A000040	Payment	D	31.07.2008, 13:43 18.12.2009, 14:25	parschalk, andreas	sdffsdff	asdf	20.30	SG	
» » UBI08A000080	Payment	D	31.07.2008, 14:05 18.12.2009, 14:25		sdffsdff_1	asdf	10.00	CM	
» » UBI08A000090	Payment	D	31.07.2008, 14:05 18.12.2009, 14:25		sdffsdff_2	asdf	10.30	CM	
UBI08A000060	Payment	D	31.07.2008, 13:44 02.02.2009, 18:50	parschalk, andreas	sdffsdff	asdf	17.52		
UBI08A000150	Payment	D	18.08.2008, 18:08 19.02.2009, 18:44	parschalk, andreas	vše v pořádku, test...		8.00		PUBLISHED
UBI08A000440	Payment	D	24.09.2008, 10:07 19.02.2009, 19:16	parschalk, andreas	Der Maler und Radie...		34.32		PUBLISHED
UBI08A000450	Payment	D	26.09.2008, 13:54 19.02.2009, 18:54	parschalk, andreas	test		10.24		PUBLISH
UBI08A000470	Payment	C	02.10.2008, 16:40 02.10.2008, 16:39	parschalk, andreas	sdffsdff		-	SG	

At the bottom of the table, there are navigation arrows and a page indicator '1 / 53'. The footer of the application includes the 'eod orderdatamanager' logo and 'Help | Log out' links.

Fig. 15: Screenshot “Finished orders” overview with new column “repository” indicating publishing status

After the manual review/validation of the data, the content will then be published to the “EOD repository” provided by UIBK (cf. 4.2 EOD Repository). This repository offers an OAI-PMH interface and exposes its data in oai_dc, which in turn will be harvested by our instance of REPOX and transformed to ESE v3.3 to be harvested by Europeana.

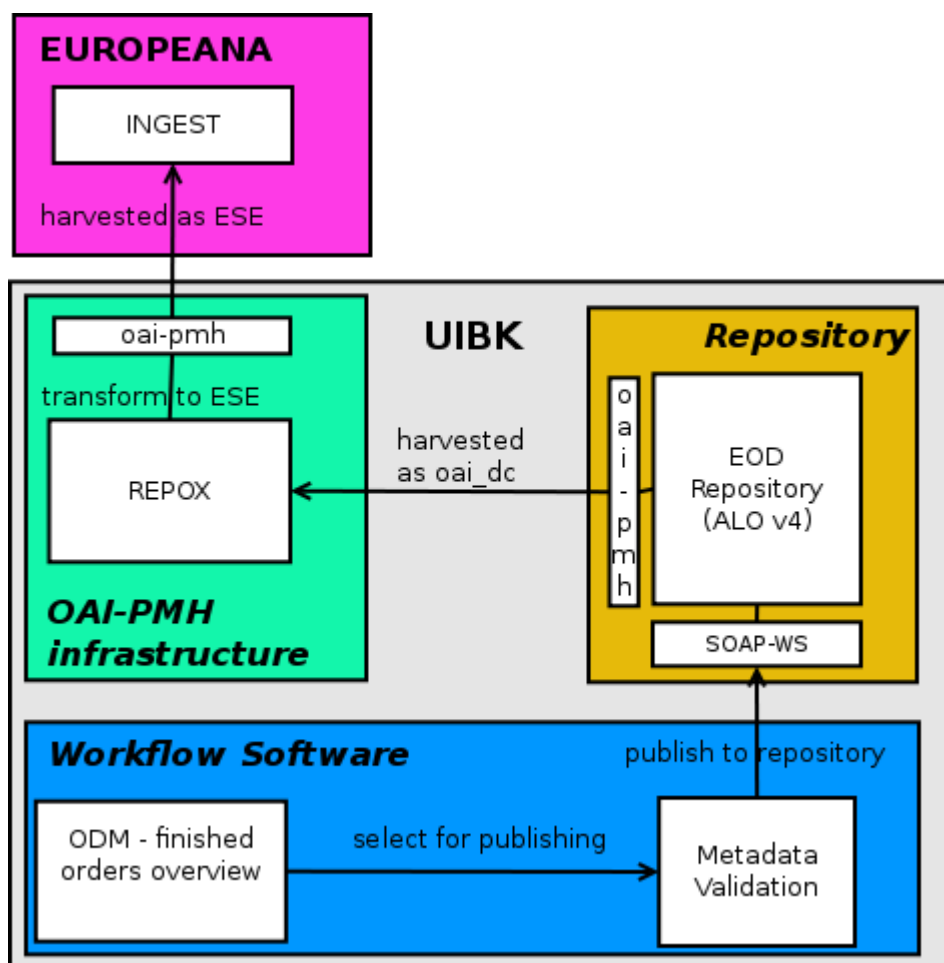


Fig. 16: Enabling content providers without harvestable repositories

4.1.4 Dealing with partners that are already aggregated by other provider(s) or are Europeana Aggregators themselves

Some of the EOD partner libraries are already submitting data via other channels to Europeana (e.g. EuropeanaLocal) or are Europeana aggregators themselves and therefore have their EOD books records already in the portal as it is the case for e.g. University Library of Vienna (data submitted via EuropeanaLocal⁷) or Royal Library of Denmark (aggregator).

In order to avoid duplicates, we can't submit this data twice. However, in order to offer future services such as "Print on Demand" it is necessary to have EOD books "tagged" in some way. Therefore, we strongly recommend to those libraries to mention "eBooks on Demand" in an appropriate field of ESE.

⁷ <http://www.europeana.eu/portal/search.html?query=ebooks+on+demand>

4.2 The EOD Repository

The creation of a harvestable (metadata) repository is central to enabling all EOD network partners to publish content to Europeana. UIBK recently deployed their fourth release of the ALO repository, a digital content and metadata repository⁸, which is now able to be used for a multi-site setup. We will use this new feature to install an EOD “skin” in ALO, where all the partners of the network may upload the digitised books with the EOD networks' digitisation workflow management software (ODM). For the current prototype, the collection “EOD network” is used.⁹

This repository (ALO) can work as a content repository as well as a metadata repository that only links to the content held offsite. Thus also partners that have a repository that is not harvestable (cf. 3 Viewer without harvestable interface) can take advantage of its OAI-PMH interface.

A SOAP based interface was created to handle the automatic transfer of digital objects and its metadata (the so called SIP package) to ALO. This interface¹⁰ is called from the ODM to conveniently publish the book to the EOD networks' repository from where it can be harvested via OAI-PMH.

⁸Cf. <http://www.literature.at> or <http://repository.uibk.ac.at>

⁹ http://www.literature.at/xims.alo?ximsurl=collections/info_1045332

¹⁰ <http://dea-atr67.uibk.ac.at/DOG/ALOService.asmx?WSDL>