VIADAT USER MANUAL

Version 1.0

Charles University, Prague, CZ

in cooperation with

Institute for Contemporary History of the Czech Academy of Sciences

and

National Film Archive

Prague
December 2019

Table of content

Viadat user manual

The following document describes basic interaction with the system from end user's point of view. There is another document concerned with how to install the system and other tasks suitable for staff (system administrators) running the system.

Viadat

The following paragraphs briefly describe interaction with the "Viadat" platform as a whole. This platform comprises of multiple modules – deposit, search, text, annotate, analyze, stat, gis – and interacts with the repository described further below. The aim of the platform is to facilitate search, information extraction, visualization and research of oral history recordings. The platform's workflow is set in such a way that forces users to mark their work as finished and publish it in the repository. Certain functionality will not be available unless the work is published. The structure of the manual sort of follows this workflow.

User Accounts

You need to be logged in to perform most of the actions. Though, at the moment all the users see all the items. To obtain a user account, you should consult your system administrator to get directions. In a fresh installation there is an administrator account "test" with password "test".

Exporting

Whenever you mark something as "ready" (as in ready to be published), you can go to the "Export" tab and click "Run Exports" (see 1 in Figure 1) this will enter all the "ready" items, upload them into the repository and change their status to "published".
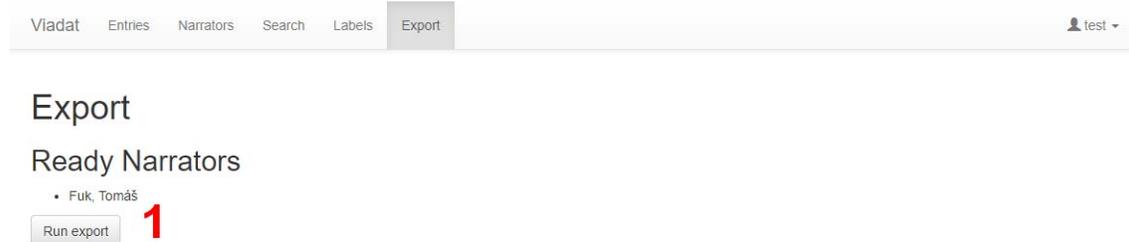

Figure 1: Run export

Create narrator

Conceptually there are two main entities that have metadata attached to them in the system. These are the narrators and the interviews. To finalize an interview you need to assign a narrator to it. Thus you need to create a narrator in the system first (unless you are adding second or any additional interviews to an already existing narrator).

After you have logged in, go to the "Narrators" tab (marked 1 in Figure 2) and click the "New narrator" button (2). Fill in the required metadata (3) and click "Create Narrator" (4). This will add a new line in the Narrators list (5). Clicking a narrators name in this list will let you edit (after clicking a value, it changes to editable field, see 1 in Figure 3) the details about the narrator and will let you change the status from private to ready (as in ready to be published).

In the current version the license is preset to "VIADAT License", consult the Adding items to the repository section of this document for what this does.

Figure 2: Create narrator

Figure 3: Edit narrator and mark as ready

Create entry

Entries sort of map to the concept of interviews, they have no other details than the name, their function is to group together multiple source items (if the interview is made of multiple files).

To create an entry go to "Entries" tab (1 in Figure 4), click "New entry" (2), fill in the entry name (3) and click "Create entry" (4). This will add a new row in the "Entries" list (5).

Figure 4: Create new entry

New source item

When you click an entry name in the entries list (5 in Figure 4) you will move to a screen where you can edit the entry name (1 in Figure 7), add new sources items (2) or remove the entry (3). Clicking "New source item" will enable upload files option (the section marked 4 will appear). In the upload file dialog (Figure 5) you choose the audio files and transcript files. See the details about these files below.
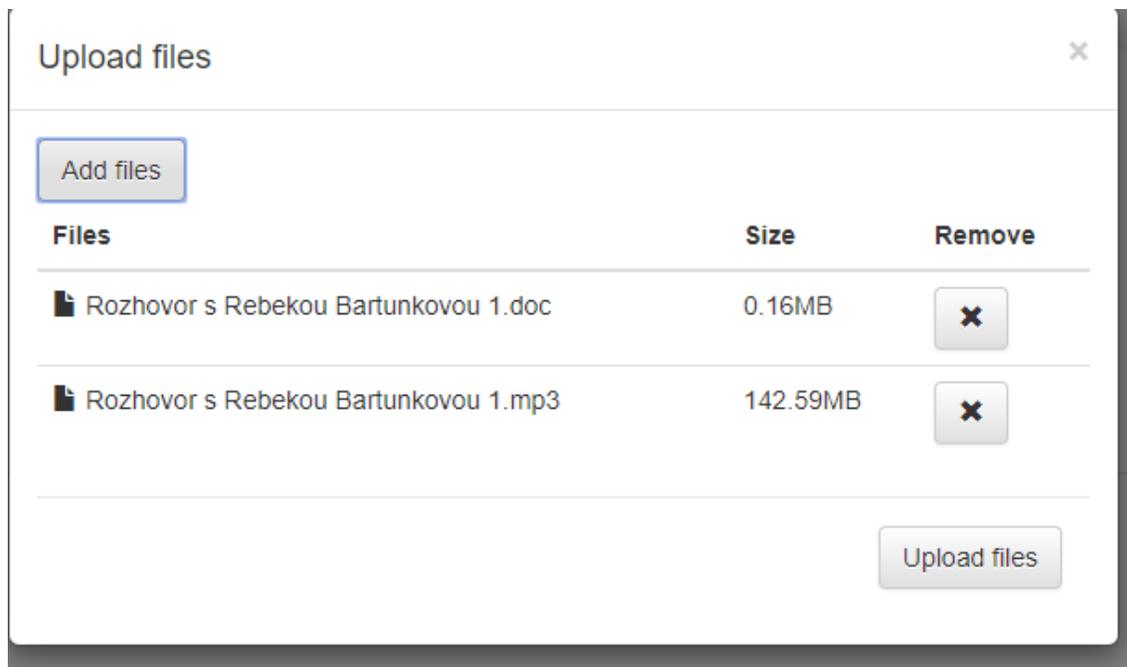
Figure 5: Upload files dialog

The "Create AT" and "Upload AT" buttons remain disabled until the source item is published. To do that, click "Metadata" link (in section 4 of Figure 7) to show an edit metadata dialog (Figure 6). Here you can try using the "Autodetect" (1) button to fill in the metadata, or fill the required details manually.

The autodetect can currently identify the title (if it is in the transcript header under "přepis rozhovoru") and the narrator (if it is in the header under "jméno a příjmení narátora/ky" and a narrator with that name was already created).

The "Update" (2) and "Set as Ready" (3) button remain disabled until all the fields contain a valid value.

The "Narrator" (4) field lets you select only from narrators that have already been published. Press the "Set as Ready" (3) button to mark the source item as ready for publishing, and then click "Update" to close the dialog.

To publish the source item the process is similar to publishing a narrator. Go to "Export" tab, under Ready sources you can see all the sources marked as ready. Clicking "Run Export" will publish all the ready items in the repository.

Figure 6: Edit metadata of a source item

File naming conventions

The allowed file types are mp3 for audio and .doc, .docx or .odt documents as the transcript corresponding to the audio file.

The name of the transcript file must be the same as the name of the corresponding audio (except for the extension). An example: `Rozhovor s Janem Sokolem.odt` and `Rozhovor s Janem Sokolem.mp3`.

An interview can be split into multiple files. In that case files need to be numbered in ascending order and the name of a file must end with its number. An example: `Rozhovor s Janem Sokolem 1.odt`, `Rozhovor s Janem Sokolem 1.mp3`, `Rozhovor s Janem Sokolem 2.odt` and `Rozhovor s Janem Sokolem 2.mp3`. The transcript

should cover only the part of audio that is in the corresponding file (ie. the audio and transcripts should be split in the same point in time).

Transcripts

Any header in the transcript file should be separated from the "proper transcript". The detector goes paragraph by paragraph and searches for a continuous section that either has a solid border-bottom on a paragraph or has a text underline and is not too long. I.e. the whole header should have a border around it, or its items should be underlined. Also the header section will end when anything longer than 20 words appears. This is important for two reasons: for auto extraction of some of the metadata, but more importantly to ease the job of the forced aligner.

The transcript text can be divided into multiple paragraphs. The first paragraph and if the speaker changes, the paragraph must start with the speaker's ID. The ID must be separated from the text by a colon. An example: "Jan Novák: Dobrý den, ...". The ID must be the same across the document.

If the speaker is unknown use "Osoba A", "Osoba B" etc. Use "Osoba ?" if it is not clear who is talking.

If the speech, or part of it, is unintelligible use "[nesrozumitelné]" either as a part of a sentence or as a whole paragraph. If used as a paragraph the rules for speaker IDs still apply.
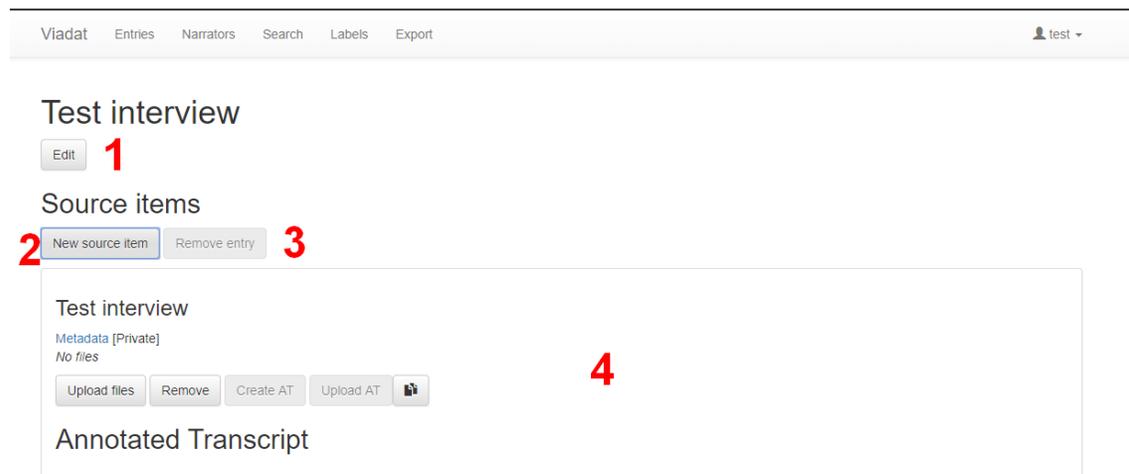

Figure 7: New source item

Create/upload annotated transcript and labels

Use the "Create AT" (Figure 8) button to align the transcript text with the sound, i.e. this lets you go to the section of audio where a given sentence is being said, you'll read about that further in the text.

When the alignment is finished new section with the annotated transcript should appear (Figure 9). In there you can inspect the alignment and add labels to the transcript (1). You can inspect the provided statistics about this transcript (2). You can download the aligned transcript and/or the labels (3). When you are done with adding labels, you can mark the transcript as ready for export (4). Going to the "Export" tab will, yet again, show the ready transcripts and clicking "Run Export" will store the annotated transcript and the labels file into the repository item that contains the audio file. After you have published the transcript you

can no longer add labels into it. It is frozen. You can clone the source item, export the clone (this will require a different signature), upload the transcript and labels to this cloned version and continue the work.
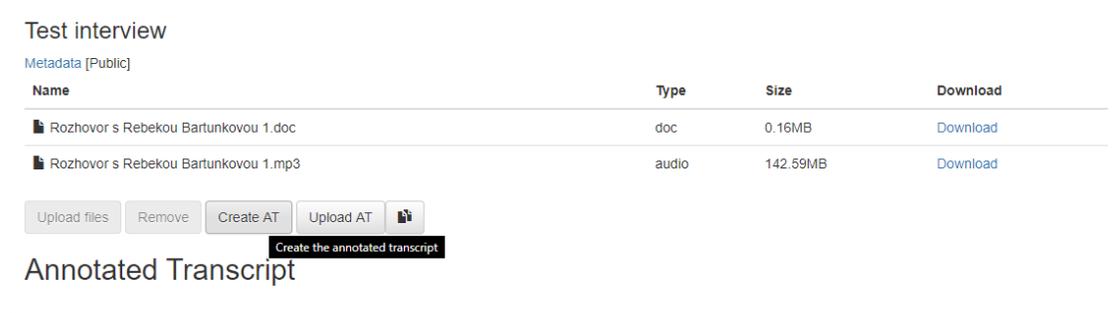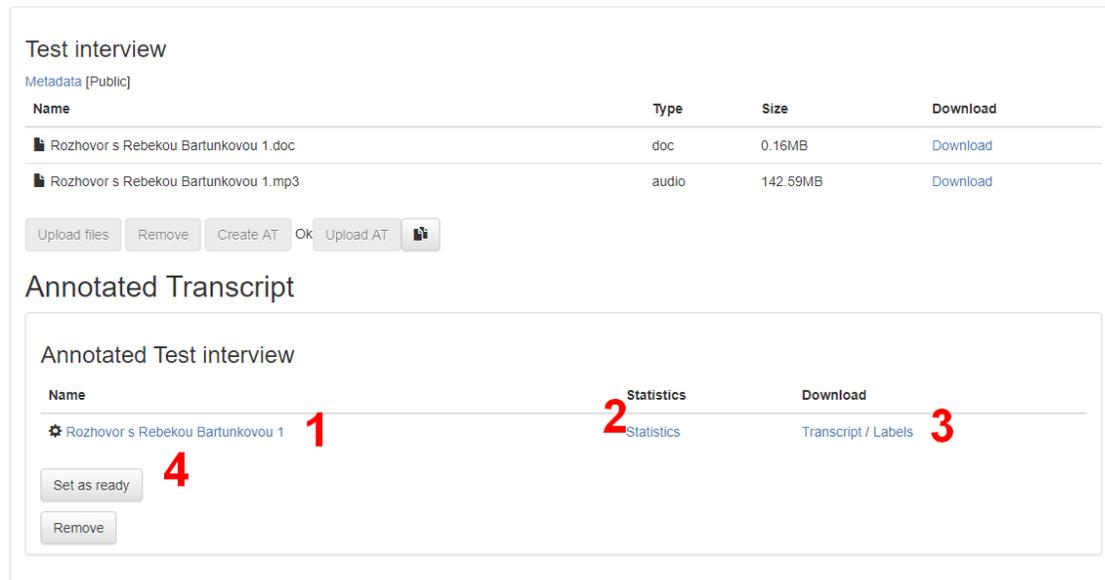

Figure 8: Create AT

Figure 9: Annotated transcript list

Audio/Labels

Clicking a transcript in the list (marked 1 in Figure 9) will take you to a page where you can play the audio or annotate/label the text. When the "Audio" tab (1) is selected you can selecte a section in the transcript (3) and listen to it (press play (4)). When you switch to the "Labels" tab (2), you can select a piece of the text, press new label and select an existing label (or create new) for the selected text. If you assign lemmas to labels, the "Show hint" checkbox will highlight the portions of text having that lemma. The lemma can be added in the "Labels" menu (1 Figure 11) after clicking the label name (2). You can also assign a

location to the label and set a from-to date for the label, these can be used for filtering the labels. It's also possible to create new labels here (3)

Rozhovor s Rebekou Bartunkovou 1

1 **2**

Audio | Labels

▶ Play | ■ Stop | Playing: 0:03 / 0:03

4

PŘEPIS ROZHOVORU	BARTŮŇKOVÁ, REBEKA (STUD-50-1)
Jméno a příjmení narátora/ky	Rebeka Bartůňková (RB)
Ročník narození	▬▬▬
Datum rozhovoru	26. 7. 2017
Místo rozhovoru	Česká televize, Kavčí Hory, Praha
Jméno a příjmení tazatele/ky	Petra Schindler-Winsten (PSW)
Délka rozhovoru	01:02:18
Počet stran přepisu	31
Projekt	Studentská generace roku 1989 v časosběrné perspektivě. Životopisná interviews po dvaceti letech (GA ČR 17-14167S).
Informovaný souhlas	ano, bez dalších ujednání

PSW

3

Takže to ještě musím na začátek uvést. Dneska je 26. července 2017, nahráváme s paní Rebekou Křižanovou - Bartůňkovou, tazatel Petra Schindler. A je to pro projekt Studenti roku 1989 po 20 letech. Já bych se možná nejdřív na začátek zeptala, jestli jste byla překvapená, že jsem se ozvali. Nebo tak vy jste asi věděla, že se na váš chystám, ale jestli vás překvapilo, že se jako pokračuje tady v tom projektu. Jestli jste byla ráda?
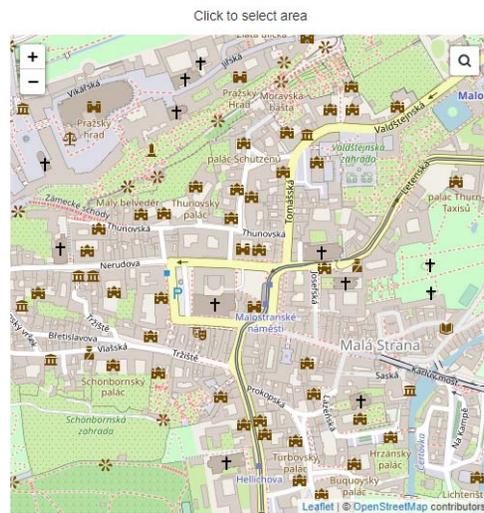
RB

Figure 10: Audio playback

Viadat　Entries　Narrators　Search　**Labels**　Export　　　　　　　　　　　　　　　　　　　　　　　　👤 test ▾

1

Labels

Filter the existing labels

By selecting an area on the map

Click to select area

Reset map filter

By setting a date filter

Show only those labels having a date range that intersects what you specify here.

Date filter

Edit the existing categories

∨ Entities Edit ▾

🏷 People **2**

Create new

New top level category **3**

Figure 11: Labels

Statistics

Clicking on the statistics link (2 in Figure 9) in the "Entry" view will show you statistics about the words used in the transcript. For example the lemma counts (see Figure 12).

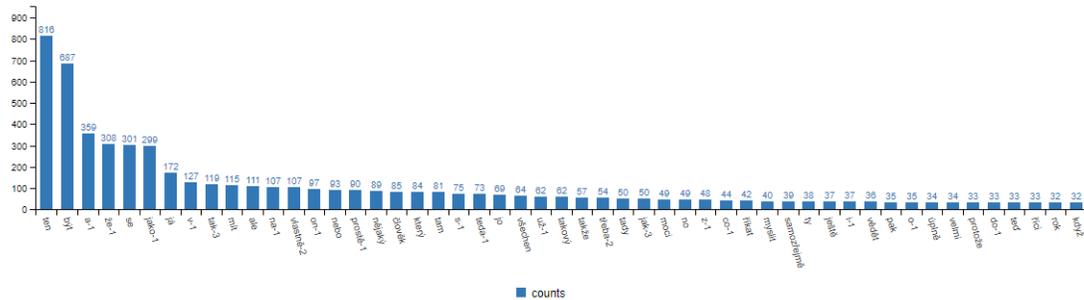Statistics of Rozhovor s Rebekou Bartunkovou 1

Lemma counts

Figure 12: Lemma counts statistics

Search

The "Search" tab (1 in Figure 13) provides a full text search through the annotated transcripts. The transcripts that match the search (the query and the transcripts are lemmatized so the matching returns more results than exact matching would) are listed (2). If you click a name of a matching transcript it will take you to the audio playback screen, where the matching words will be highlighted in the text.

Figure 13: Search

Repository

The following paragraphs describe the repository system. The point of a repository system is to keep data safe from modifications and to simplify discovery of data by searching/browsing through associated metadata.

User Accounts

To obtain a user account, you should consult your system administrator to get directions. In a fresh installation there is an administrator account "dspace@lindat.cz" with password "dspace".

Search/Browse

The following paragraphs briefly present the user interface for browsing and searching items (narrators and interviews) in the repository.

The Homepage

Search field (marked 1 in Figure 14) is used for full-text search. All metadata items are searched. Special syntax can be used to search in specific fields[1], but you can use the predefined filters and facets to further refine the search. When you press the button, you will land on the search page (described further below) where the items matching the search criteria are shown. If you do not enter any text, all records will be displayed.
The link "Advanced Search" (2) will also take you to the search page. In this case you'll be shown an additional form for filtering on particular metadata fields.
The selected facets (3), a subset of the facets that are available on the search page, namely Project, Topic, Profession, Keywords and Period, once again lead to the search page. This time only the records having the given facet value are shown. For example, all interviews from the selected period. The facets appear only if they can reduce the number of results, this behavior is the same on the search page.
The homepage also includes newly added and frequently visited items (narrators and interviews (4)).
Records in the repository can be browsed (5) in various orderings: issued date, language and names (including aliases).
When you click on the name of an item (whether on homepage, search or when browsing), the item view page opens (described further below).
If a user is logged and has sufficient privileges, the right hand side menu (6) shows a context section. This is where you can add new narrators to the repository, and if you are on narrator's item-view pages (see below) this is where you can add new interviews to the narrator.

---

[1] But this is beyond this introduction, please see the SOLR syntax documentation at
https://cwiki.apache.org/confluence/display/solr/The+Standard+Query+Parser

Repozitář Centra orální historie ÚSD AV ČR

COH   ÚSD

1 Q [   ] Hledat

2 Rozšířené hledání

Projekt	Téma	Profese	Klíčová slova - rozhovor	Detailní klíčová slova	Období
eu-fp-10 (4)	tema_klíč5 (10)	prof_slovo8 (7)	int_klíč1 (10)	det_slovo7 (12)	Machine Age (6)
eu-fp-2 (4)	tema_slovo3 (7)	prof_klíčové slovo9 (6)	int_klíč5 (8)	det_klíč6 (8)	Bosnian War (4)
eu-fp-12 (3)	tema_slovo7 (7)	prof_klíč9 (5)	int_klíč4 (6)	det_klíč7 (8)	Interwar Period (4)
eu-fp-14 (3)	tema_klíč10 (6)	prof_slovo1 (5)	int_klíčové slovo6 (6)	det_klíč3 (7)	The Nineties (4)
eu-fp-4 (3)	tema_klíč1 (5)	prof_klíč2 (4)	int_klíčové slovo7 (6)	det_klíč4 (7)	War on Terrorism (4)
... zobrazit další	... zobrazit další	... zobrazit další	... zobrazit další	... zobrazit další	... zobrazit další

3

Nově přidané

Interview   VIADAT

title9205

Popis:
Dolorem modi neque tempora adipisci amet est etincidunt. Quisquam ipsum velit consectetur. Magnam quiquia etincidunt numquam. Etincidunt eius numquam consectetur modi eius ipsum. Sed velit magnam aliquam est amet ipsum.

🔗 Rozhovor obsahuje 1 soubor (0 bajtů).

Interview   VIADAT

title57

Popis:
Aliquam sed magnam velit quaerat quiquia modi. Ipsum dolor dolor numquam tempora numquam consectetur sit. Neque dolor velit porro amet. Eius est adipisci ut ut eius. Porro magnam non modi.

🔗 Rozhovor obsahuje 2 souborů (0 bajtů).

Interview   VIADAT

title3540

Popis:
Dolore etincidunt non neque eius magnam eius. Numquam dolor magnam dolorem. Est voluptatem dolor ipsum modi tempora. Aliquam amet dolorem tempora eius adipisci. Neque consectetur modi porro. Aliquam tempora numquam modi ...

🔗 Rozhovor obsahuje 2 souborů (0 bajtů).

4

Nejnavštěvovanější záznamy

Za poslední týden

Narrator   VIADAT

Svobodová, Lucie

Alias:
Novotná, Zdeňka ; Kučerová, Markéta ; Benešová, Jarmila ; Sedláčková, Jarmila ; Hájková, Jitka
Klíčová slova:
prof_klíč6 ; tema_klíč5 ; tema_slovo7

👤 Počet rozhovorů s tímto narátorem: 3 🔗 Narátor poskytl 1 soubor (0 bajtů).

❓ Kde začít?

DEPOSIT   CITE

◎ Procházet

› Vše v repozitáři

👤 Můj účet

🔓 Odhlásit se
👤 Profil
⬆ Příspěvky

☰ Kontext

📄 Vytvořit komunitu
› Přidat narátora

5

6

⚙ Správa

🔧 Ovládací panel
👁 Řízení přístupu
› Správa obsahu
▦ Registry
🔗 Collections & Communities
📊 Statistiky
📱 Úlohy správy
✎ Licence
👉 Handles
› External Handles

ℹ Obecné informace

⬆ Nahrát
💬 Cituj
🔄 Stavy záznamu
❓ FAQ
ℹ O repozitáři
✉ Poradna

📊 Statistiky

› Zobrazit statistiky využívání
› Zobrazit statistiky vyhledávání
› Zobrazit statistiky workflow
📊 Google Analytics

📡 RSS Feed

📡 RSS 1.0
📡 RSS 2.0
📡 Atom

Figure 14: The homepage

The search page

The search page (Figure 15) again contains a search field (1) and facets (4). The facets are divided into 3 levels depending on the type of metadata on which they filter (narrator, interview and technical details).

There are extended search filters (2), previews of items that match the search criteria (5) and a switch (3) for the type of results (narrators/interviews) shown.

While facets work with exact values, filters can be used to enter queries like: "the name contains 'Pavel'". When entering a filter value, auto-completion is offered.

The 3 levels of facets and the result type switch makes it convenient to show, for example, all the narrators having an interview in English or, on the other hand, show all interviews where the narrator is a doctor (has a doctor's degree).

Facet remains hidden if its choice would not reduce the number of results displayed.

Facets over dates are dynamic in the sense that they create several roughly equal groups, and the from-to range of these groups changes depending on the current search results.

The item-view page

Individual record pages may be displayed in simple or full mode. In both modes, there is always a refbox ("how to cite", 1 in Figure 16), metadata (2), the files (3) associated with the given record, depending on the type of item these are either the files with the interview or files supplied by the narrator (if any). If the item describes a narrator, both views also show and link to interviews with this narrator (4).

Simple mode differs from full mode in how and which metadata are displayed. The simple item view serves as sort of a summary of a record, only a selection (see Figure 16) of metadata is shown. In the full mode all the metadata might be displayed (see Figure 17) including the generated "administrative" metadata. Whether those will be shown depends on the privileges of the current user.

In Figure 16 you can also see the context menu (5) now has "Add Interview" link.

Figure 15: The search page

Figure 16 Item view page displaying a narrator in simple mode

Zobrazit minimální záznam

❝ Pro citaci použijte následující text, nebo export do připraveného formátu:

Narrator: Zeman, Michal, Oral History Center, http://hdl.handle.net/123456789/xyz-00040939.

BIBTEX CMDI

◁ Sdílet:

dc.date.accessioned	2020-01-08T10:29:33Z
dc.date.available	2020-01-08T10:29:33Z
dc.identifier	xyz-00040939
dc.identifier.uri	http://hdl.handle.net/123456789/xyz-00040939
dc.relation.haspart	http://hdl.handle.net/123456789/xyz-04443065
dc.relation.haspart	http://hdl.handle.net/123456789/xyz-08045749
dc.rights	VIADAT License
dc.rights.uri	https://ufal.mff.cuni.cz/grants/viadat/license
dc.rights.label	RES
dc.title	Zeman, Michal
dc.type	narrator
has.files	yes
branding	VIADAT
files.size	0
files.count	1
viadat.narrator.gender	muž
viadat.narrator.birthdate	1946-11-05
viadat.narrator.alias	Horák, Michal
viadat.narrator.alias	Zeman, Michal
viadat.narrator.alias	Černý, Miloslav
viadat.narrator.degree	Ing.
viadat.narrator.degree	Ing.
viadat.narrator.degree	Ph.d.
viadat.narrator.keywordsProfession	prof_klíčové slovo4
viadat.narrator.keywordsProfession	prof_slovo1
viadat.narrator.keywordsTopic	tema_klíčové slovo5
viadat.narrator.keywordsTopic	tema_klíčové slovo8
viadat.narrator.keywordsTopic	tema_klíčové slovo8
viadat.narrator.keywordsTopic	tema_klíčové slovo4
viadat.narrator.keywordsTopic	tema_klíčové slovo1
viadat.narrator.consent	Ano
viadat.narrator.contact	narrator7296
viadat.narrator.note	narrator8604
viadat.narrator.project	eu-fp-4
viadat.narrator.outputs	narrator8246
viadat.narrator.materials	narrator7514

🔗 Soubory poskytnuté narátorem

Název	photo.jpg

Licenční kategorie: Restricted Use
Licence: VIADAT License

Zobrazit minimální záznam

🎤 Rozhovory s tímto narátorem

Interview   VIADAT

title57

Klíčová slova:
int_klíčové slovo6 ; int_klíč9 ; int_klíčové slovo1 ; int_klíčové slovo2 ; int_klíčové slovo8 ; det_slovo10 ; det_klíčové slovo3 ; det_slovo1

Popis:
Aliquam sed magnam velit quaerat quiquia modi. Ipsum dolor dolor numquam tempora numquam consectetur sit. Neque dolor velit porro amet. Eius est adipisci ut ut eius. Porro magnam non modi.

🔗 Rozhovor obsahuje 2 souborů (0 bajtů).

Interview   VIADAT

title9205

Klíčová slova:
int_klíč4 ; det_klíčové slovo2 ; det_klíčové slovo8 ; det_klíč4 ; det_klíčové slovo2 ; det_klíč7

Popis:
Dolorem modi neque tempora adipisci amet est etincidunt. Quisquam ipsum velit consectetur. Magnam quiquia etincidunt numquam. Etincidunt eius numquam consectetur modi eius ipsum. Sed velit magnam aliquam est amet ipsum.

🔗 Rozhovor obsahuje 1 soubor (0 bajtů).

Figure 17: Item view page displaying a narrator in full mode

Adding items to the repository

The process for adding narrators and interviews follows the same pattern, you are taken through a series of steps (pages) where you fill in the required information, upload the files etc. What is different between the two is the information you are asked to fill in.
You can upload files either from your personal computer (marked 1 in Figure 18), that's the one with your browser open, or you can specify a URL (2). The URL can be anything on the Internet, but also a path on the server, the machine where the repository is running. In both cases, you choose the type of file (interview file, transcript, material, informed consent). There is a file size limit of 2GB for files uploaded from personal computers (1).
Even though you can use the "Submissions" menu to create new records (the same way as in stock DSpace), you should generally avoid that and use the "Add Narrator"/"Add Interview" links that appear in the context menu (see The item-view page above). These links prefill certain kind of information that you otherwise would have to fill in manually. Use the "Submissions" menu to continue creating a narrator/interview if you have saved your progress for later.

Item submission

Figure 18: File upload step

Adding a narrator

When a user is logged in and if he has appropriate access rights, the context menu in the column on the right hand side will show an "Add Narrator" link. Clicking that will start a new submission in the "narrators" collection.

The first "Basic Info" step gathers information about the narrator – name, aliases, gender, birthdate, id and various other specified in the metadata schema. The id will be used as the

"suffix" part of a handle[2] (if your deployment uses the handle system), that is why it must be unique (not used for any other narrator, interview or other DSpace object).

In the next step you can upload files provided by the narrator, for example photos, or the informed consent.

If you happen to upload any files a "Restricted Use" license is attached to the item. This servers multiple purposes, one of those is limiting who can access the files later on, by requiring users to login in (and you can further refine the privileges based on login).

Adding an interview with a narrator

When you are viewing a narrator's page (see The item-view page) the context menu on your right should show an "Add interview" link, clicking that will start a new submission in the "interviews" collections and fill in metadata required to link the narrator with the interview. First, you are required to upload files. Details about uploading files mentioned above still apply, the same goes for the license. If you click edit near the uploaded file, you can fill in some more details about the particular file.

In the next step, you provide information about the interview such as the language, date, length etc. What's mentioned about ids in the narrator submission applies here too. The first field called "Title" should be something convenient that would let you discern the interviews in the search results.

---

[2] To avoid assigning multiple ids to one "entity", or to at least provide straight forward mapping between these ids.