Caricamento tesi su BNCF

Procedura di invio tesi al nuovo servizio di deposito legale.

A differenza della versione precedente nella quale BNCF accedeva ai vari servizi OAI-PMH degli atenei recuperando le nuove tesi il nuovo workflow prevede che ogni ateneo proceda con l'invio dei propri dati verso il repository mediante l'invocazione di specifici servizi REST. Ogni client deve quindi richiedere le credenziali per l'invio che identificheranno l'ente inviante all'interno del sistema di deposito legale e che abiliteranno alla creazione/modifica dati esclusivamente per le proprie tesi.

Possiamo distinguere 3 differenti workflow a seconda che si voglia aggiungere una nuova tesi, aggiornarne una già presente o eliminarne una precedentemente inviata.

Invio nuova tesi

Step 1 - Recupero identificativo collection di appartenenza

Come prima cosa occorre recuperare l'identificativo della collection di appartenenza (tipicamente questa rappresenta il contenitore unico di tutte le tesi dell'ente inviante) che dovrà poi essere usato successivamente in fase di invio tesi.

Endpoint: https://tesidottorato.depositolegale.it/rest/collections

Method: GET

Autenticazione: Basic

Parametri:

- authorized: valorizzare a true (**obbligatorio**)

Descrizione: In caso di chiamata corretta nel campo uuid della risposta viene indicato l'identificativo della collection

Esempio CURL:

curl --location 'https://test.bncf.unitesi.cineca.it/rest/collections' \
--header 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxx' \

Step 2 - Invio metadati tesi

Una volta recuperato l'identificativo della collection è possibile procedere con l'invio di tutti i metadati della tesi. L'operazione comporterà la creazione della tesi sul repository remoto

Endpoint: https://tesidottorato.depositolegale.it/rest/collections/<uuid_collection>/items

Method: POST

Autenticazione: Basic

Descrizione: Creazione di una nuova tesi contenente i metadati elencati nel body del messaggio. Il campo uuid_collection nell'URL contiene l'identificativo della collezione recuperato allo step 1.

Body: Oggetto in formato JSON contenente i metadati della tesi es:

```
{
   "handle":"123456789/1234",
   "submitter":"XXXXXX",
   "metadata":[
      {"key": "dc.title","value": "test"},
      {"key": "dc.authority.people","value": "Mario, ROSSI","language": "it","authority": "RSSMRA70B20A944K"}
   ]
}
```

Di seguito il dettaglio della struttura dell'oggetto da inserire nel body del messaggio:

- handle: campo opzionale contenente un identificativo unico sul sistema inviante. Per gli invii da IRIS viene utilizzato l'handle della tesi registrato su handle.net. Il campo non ha controllo di formato quindi applicativi di terze parti possono inserire valori custom. Se presente viene salvato nel metadato dc.identifier.remoteid. Se il dato è presente la creazione del nuovo item viene interrotta qualora venga trovato nel sistema un altro item pubblico con lo stesso valore di identificativo remoto;
- submitter: campo obbligatorio contenente il nome dell'utente che carica il dato (utilizzare sempre lo username fornito da Cineca per l'accesso a BNCF);
- metadata: array contenente la lista dei metadati caricabili su BNCF. I metadati possono prevedere una authority (corrispondente al codice fiscale) utilizzata dal sistema per discriminare tra le omonimie. Per quanto riguarda i metadati senza authority occorre fornire la chiave e il valore del metadato, es: {"key": "dc.title","value": "test"} mentre per i metadati con authority la struttura sarà {"key": "dc.authority.people","value": "Mario, ROSSI","language": "it","authority": "RSSMRA70B20A944K"}. Nei casi di metadati come authority indicare sempre "it" nel campo language. I valori supportati da BNCF per il campo key sono i seguenti:
    - dc.authority.people (obbligatorio) - Studente (formato consigliato "nome, cognome") - L'authority è obbligatoria e deve contenere il codice fiscale;
    - dc.relation.matricola - Matricola;
    - dc.relation.department - Dipartimento di afferenza;
    - dc.relation.course - Corso di studio (metadato non ripetibile);
    - dc.date.issued (obbligatorio) - Data di pubblicazione (metadato non ripetibile);
    - dc.coverage.academicyear - Anno accademico (metadato non ripetibile);
    - dc.coverage.academiccycle - Ciclo accademico (metadato non ripetibile);
    - dc.authority.academicField2000 - SSD (antecedenti Maggio 2024), se valorizzato l'authority è obbligatoria e deve contenere il codice SSD;
    - dc.authority.academicField2024 - SSD (nuovi SSD 2024) , se valorizzato l'authority è obbligatoria e deve contenere il codice SSD;
    - dc.title (obbligatorio) - Titolo (metadato non ripetibile);
    - dc.title.eng - Titolo in lingua inglese (metadato non ripetibile);
    - dc.language.iso - Lingua - usare codici a 3 caratteri secondo lo standard ISO-639;
    - dc.description.abstractita - Abstract in italiano (metadato non ripetibile);
    - dc.description.abstracteng - Abstract in inglese (metadato non ripetibile);
    - dc.subject.keyword - Keywords;
    - dc.authority.advisor - Relatore (formato consigliato "nome, cognome") - ripetibile. Possibilità di indicare l'authority (codice fiscale);
    - dc.authority.otherpeople - Correlatore, Controrelatore, Co-Supervisor, Co-Tutor o Coordinatori (formato consigliato "nome, cognome") - ripetibile. Possibilità di indicare l'authority (codice fiscale);
    - dc.identifier.url - URL (metadato non ripetibile);
    - dc.publisher.name - Nome editore (metadato non ripetibile);
    - dc.publisher.place - Città editore (metadato non ripetibile);
    - dc.relation.numberofpages - Numero di pagine (metadato non ripetibile);
    - dc.type.referee - Referee - Valori possibili: "Si", "No", "" (metadato non ripetibile);
    - dc.contributor.country - Nazionalità autore - usare codici a 3 caratteri secondo lo standard ISO-3166 (metadato non ripetibile);
    - dc.description.europaeus - Certificato Doctor Europaeus - Valori possibili: "Si", "No", "" (metadato non ripetibile);
    - dc.description.cotutela - Tesi di dottorato in cotutela - Valori possibili: "Si", "No", "" (metadato non ripetibile);
    - dc.description.cotutelacountry - Paese di cotutela - usare codici a 3 caratteri secondo lo standard ISO-3166 (metadato non ripetibile);
    - dc.description.unipartner - Università straniera partner (solo se cotutela, metadato non ripetibile);
    - dc.identifier.nbn - Identificativo NBN (metadato non ripetibile);

In caso di errori di validazione del contenuto del body il sistema restituisce un errore indicando il metadato errato o la causa dell'errore.

In caso di successo il messaggio di risposta del servizio contiene l'identificativo della tesi appena creata da utilizzare per l'eventuale step 3.

ESEMPIO CURL:

```
curl --location 'https://test.bncf.unitesi.cineca.it/rest/collections/876ba4ab-c292-4e8b-91b3-51e729adff0a/items' \
--header 'Content-Type: application/json' \
--header 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxx' \
--data '{


"handle":"123456789/1234",

"submitter":"XXXXXX",

"metadata":[

{"key": "dc.title","value": "test"},

{"key": "dc.authority.people","value": "Mario, ROSSI","language": "it","authority": "RSSMRA70B20A944K"}

]

}'
```

Step 3 - Invio allegati tesi

Dopo aver creato l'item è eventualmente possibile aggiungere allegati di dimensione massima 800MB. Le richieste inviate ai servizi attraversano un filtro Cloudflare che rifiuta qualsiasi richiesta di dimensione superiore a 100MB. Sono quindi previste due modalità di invio:

- invio standard (descritto di seguito): per allegati di dimensione fino a 90MB, richiamando più volte il servizio seguente;
- invio a blocchi (descritto nello Step 3 bis): per allegati di dimensione superiore a 90MB.

ATTENZIONE: l'invio con modalità standard di un allegato di dimensione superiore al limite restituisce l'errore HTTP 413 "Payload Too Large" (pagina HTML con la dicitura "cloudflare"). In questo caso la richiesta non raggiunge il repository e occorre utilizzare l'invio a blocchi.

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/bitstreams

<uuid_item> deve essere valorizzato con l'UUID restituito nella chiamata al servizio dello Step 2

Method: POST

Content-type: MULTIPART/FORM-DATA (i parametri "**access**", "**name**", "**description**", "**date**", "**license**" devono essere indicato nel URL della chiamata, anche detto query string, e non nei FORM-DATA)

Autenticazione: Basic

Parametri:

- access: policy di accesso. Accetta il valore "openAccess" per i bitstream open access, "embargo" per i bitstream con data di embargo e "archiveadmin" per i bitstream privati (**obbligatorio**);
- date: data di embargo nel formato yyyy-mm-dd (**obbligatorio solo se access valorizzato a embargo**);
- name: nome del bitstream (**obbligatorio**);
- file: allegare file da inviare (**obbligatorio**);
- description: descrizione file;
- license: licenza dell'allegato. Il parametro è facoltativo. Qualora non venga inviato o venisse utilizzato un valore non facente parte della lista seguente verrà attribuita da default una licenza denominata "Tutti i diritti riservati". Il parametro accetta i seguenti valori:
    - publicdomain: Pubblico Dominio;
    - by: Creative commons Attribution 4.0 International;
    - by-nd: Creative commons Attribution-NoDerivatives 4.0 International;
    - by-sa: Creative commons Attribution-ShareAlike 4.0 International;
    - by-nc: Creative commons Attribution-NonCommercial 4.0 International;
    - by-nc-sa: Creative commons Attribution-NonCommercial-ShareAlike 4.0 International;
    - by-nc-nd: Creative commons Attribution-NonCommercial-NoDerivatives 4.0 International;
    - IRIS.PRI02: Tutti i diritti riservati.

ATTENZIONE **i parametri sono case sensitive**

Descrizione: Creazione di un nuovo bitstream. uuid_item rappresenta l'identificativo del nuovo item creato su BNCF restituito dal servizio allo step precedente

Body: File da allegare

ESEMPIO CURL:

curl --location 'https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/bitstreams?access=openAccess&license=publicdomain&name=file.pdf' \

--header 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx' \

--header 'Content-Type: multipart/form-data' \

--form 'file=@"/path/to/file.pdf"'

Step 3 bis - Invio allegati di dimensione superiore a 90MB (invio a blocchi)

Per gli allegati di dimensione superiore a 90MB il file deve essere suddiviso in blocchi di dimensione massima 90MB. Ogni blocco viene inviato con una chiamata separata e al termine una chiamata di completamento ricompone il file e crea l'allegato, con le stesse regole e gli stessi controlli dell'invio standard. La procedura prevede le seguenti fasi:

- Fase 1: inizializzazione dell'invio;
- Fase 2: invio di tutti i blocchi;
- Fase 3: verifica dei blocchi ricevuti (opzionale);
- Fase 4: completamento dell'invio;
- Fase 5: annullamento dell'invio (opzionale).

<uuid_item> deve essere valorizzato con l'UUID restituito nella chiamata al servizio dello Step 2.

Fase 1 - Inizializzazione

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/bitstreams/chunked/init

Method: POST

Autenticazione: Basic

Parametri (da indicare nel URL della chiamata, anche detto query string):

- name: nome del file;
- size: dimensione complessiva del file in byte;
- checksum: checksum SHA-256 del file, in formato esadecimale. Il parametro è facoltativo ma consigliato: se indicato, al completamento il sistema verifica che il file ricomposto corrisponda al checksum e in caso contrario rifiuta l'operazione.

Descrizione: Crea una sessione di invio. La risposta, in formato JSON, contiene il campo uploadId, da utilizzare in tutte le fasi successive, e il campo chunkSize, che indica la dimensione massima in byte di ciascun blocco (90MB, pari a 94371840 byte).

ESEMPIO CURL:

```
curl --location --request POST 'https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/bitstreams/chunked/init?name=file.
pdf&size=591379231&checksum=d9dcbfa3c6144b18c41b3b73b58fb3f277fb99ebdd65be36345d8ee6283febe0'\
  --header 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx' \
  --header 'Accept: application/json' \
  --header 'Content-Length: 0'
```

Fase 2 - Invio dei blocchi

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/bitstreams/chunked/<uploadId>/<indice>

Method: PUT

Content-type: application/octet-stream (il body contiene direttamente i byte del blocco, non si utilizza MULTIPART/FORM-DATA)

Autenticazione: Basic

Descrizione: Invio di un singolo blocco. <uploadId> è il valore restituito dalla Fase 1. <indice> è il numero progressivo del blocco, a partire da 0 (primo blocco = 0, secondo blocco = 1, ecc.). Ogni blocco deve avere dimensione massima 90MB. Un blocco già inviato può essere inviato nuovamente (ad esempio dopo un errore di rete): il contenuto precedente viene sostituito. La chiamata deve essere ripetuta per tutti i blocchi del file.

ATTENZIONE: gli indici devono essere consecutivi a partire da 0, senza interruzioni, altrimenti il completamento dell'invio (Fase 4) viene rifiutato.

ESEMPIO CURL + suddivisione del file e invio di tutti i blocchi da shell Linux:

```
sha256sum file.pdf

split -b 90M -d -a 4 file.pdf blocco_

i=0
for f in blocco_*; do
  curl --location --request PUT "https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/bitstreams/chunked/<uploadId>/$i" \
  --header 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx' \
  --header 'Content-Type: application/octet-stream' \
  --data-binary "@$f"
  i=$((i+1))
done
```

Fase 3 - Verifica dei blocchi ricevuti (opzionale)

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/bitstreams/chunked/<uploadId>

Method: GET

Autenticazione: Basic

Descrizione: Restituisce lo stato della sessione. Il campo received contiene l'elenco degli indici dei blocchi già ricevuti. Il servizio permette di riprendere un invio interrotto inviando soltanto i blocchi mancanti.

ESEMPIO CURL:

```
curl --location 'https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/bitstreams/chunked/<uploadId>' \
--header 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx' \
--header 'Accept: application/json'
```

Fase 4 - Completamento

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/bitstreams/chunked/<uploadId>/complete

Method: POST

Autenticazione: Basic

Parametri: gli stessi dello Step 3 ("access", "name", "description", "date", "license"), con le stesse regole di obbligatorietà, da indicare nel URL della chiamata. La chiamata non ha body.

Descrizione: Ricompone il file a partire dai blocchi ricevuti e crea il bitstream. Se mancano dei blocchi il servizio restituisce un errore che ne indica gli indici; se in Fase 1 è stato indicato il checksum e il file ricomposto non corrisponde, l'operazione viene rifiutata. In caso di successo la risposta ha lo stesso formato dello Step 3 e la sessione di invio, con i relativi blocchi temporanei, viene eliminata. Per file di grandi dimensioni l'operazione può richiedere alcuni secondi: configurare nel client un timeout adeguato.

ATTENZIONE: se la chiamata di completamento termina per timeout, prima di ripeterla verificare che l'allegato non sia già presente nella tesi, per evitare di crearne un duplicato.

ESEMPIO CURL:

curl --location --request POST 'https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/bitstreams/chunked/<uploadId>
/complete?access=openAccess&license=publicdomain&name=file.pdf' \
--header 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx' \
--header 'Accept: application/json' \
--header 'Content-Length: 0'

Fase 5 - Annullamento (opzionale)

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/bitstreams/chunked/<uploadId>

Method: DELETE

Autenticazione: Basic

Descrizione: Annulla l'invio ed elimina i blocchi già caricati. Le sessioni non completate e inutilizzate per un periodo prolungato vengono comunque eliminate automaticamente dal sistema.

ESEMPIO CURL:

curl --location --request DELETE 'https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/bitstreams/chunked/<uploadId>' \
--header 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx'

Step 4 - Pubblicazione tesi

Fino al completamento con successo dello step precedente la tesi e gli eventuali allegati risultano non visibili nell'area pubblica dell'applicazione. Lo step corrente rende la tesi visibile all'esterno

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/workflowSetStateArchive

Method: PUT

Autenticazione: Basic

Descrizione: Pubblicazione della tesi. uuid_item rappresenta l'identificativo del nuovo item creato su BNCF restituito dal servizio allo step 2

ESEMPIO CURL:

curl -X PUT https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/workflowSetStateArchive \
  -H 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx'

Modifica tesi esistente

Step 1 - Cancellazione metadati

Esegue la cancellazione di tutti i metadati di una tesi già esistente. Se la tesi risultava pubblica viene nascosta

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/metadata

Method: DELETE

Autenticazione: Basic

Descrizione: Cancellazione di tutti i metadati dell'item. uuid_item rappresenta l'identificativo dell'item su BNCF precedentemente creato

ESEMPIO CURL:

curl --location --request DELETE 'https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/metadata' \
--header 'Authorization: Basic XXXXXXXXXXXXXXXXXXXXXXXXX' \

Step 2 - Invio metadati tesi

Una volta cancellati i vecchi metadati è possibile procedere con l'invio dei nuovi.

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/metadataItem

Method: POST

Autenticazione: Basic

Descrizione: Creazione metadati per un item esistente.

Body: Oggetto in formato JSON contenente i metadati della tesi secondo le specifiche fornite allo step 2 del workflow di creazione tesi

ESEMPIO CURL:

```
curl --location 'https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/metadataItem' \
--header 'Authorization: Basic xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx' \
--header 'Content-Type: application/json' \
--data '{

    "uuid":"e2159bb3-b787-454f-ac34-a465996fdc84",

    "handle":"123456789/1234",

    "submitter":"unimi-ws",

    "metadata":[

        {"key": "dc.title","value": "test"},

        {"key": "dc.authority.people","value": "Mario, ROSSI","language": "it","authority": "RSSMRA70B20A944K"},

        {"key": "dc.relation.course","value": "un corso studi","language": "it","authority": "RSSMRA70B20A944K"},

        {"key": "dc.coverage.academicyear","value": "2020","language": "it","authority": "RSSMRA70B20A944K"},

        {"key": "dc.coverage.academiccycle","value": "XXV","language": "it","authority": "RSSMRA70B20A944K"},

        {"key": "dc.authority.advisor","value": "Relatore","language": "it"},

        {"key": "dc.language.iso","value": "ira","language": "it","authority": "RSSMRA70B20A944K"},

        {"key": "dc.authority.advisor","value": "Relatore authority","language": "it","authority": "RSSMRA70B20A944K"},

        {"key": "dc.date.issued","value": "2023-06-09","language": "it","authority": "RSSMRA70B20A944K"},

        {"key": "dc.authority.otherpeople","value": "Correlatore senza authority","language": "it"},

        {"key" : "dc.identifier.nbn","value" : "urn:nbn:it:xxxx-00000"  },

        {"key": "dc.authority.otherpeople","value": "Correlatore","language": "it","authority": "RSSMRA70B20A944K"}


    ]

}'
```

Step 3 - Cancellazione bitstream

Esegue la cancellazione di tutti i bitstream di una tesi già esistente.

Endpoint: https://tesidottorato.depositolegale.it/rest/items/<uuid_item>/bitstreams

Method: DELETE

Autenticazione: Basic

Descrizione: Cancellazione di tutti i bitstream dell'item. uuid_item rappresenta l'identificativo dell'item su BNCF precedentemente creato

ESEMPIO CURL:

```
curl --location --request DELETE 'https://test.bncf.unitesi.cineca.it/rest/items/e2159bb3-b787-454f-ac34-a465996fdc84/bitstreams' \
--header 'Authorization: Basic XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX' \
```

Step 4 - Invio allegati tesi

Come step 3 del workflow di creazione nuova tesi

Step 5 - Pubblicazione tesi

Come step 4 del workflow di creazione nuova tesi

Recupero tesi esistente

Servizio utilizzabile per il recupero di una tesi precedentemente inviata noto l'identificativo univoco locale (trasmesso a BNCF nel campo handle in fase di creazione della tesi) .

Endpoint: https://tesidottorato.depositolegale.it/rest/items/find-by-metadata-field

Method: POST

Autenticazione: Basic

Descrizione: Ricerca di una tesi noto il suo identificativo locale

ESEMPIO CURL:

curl -X POST https://test.bncf.unitesi.cineca.it/rest/items/find-by-metadata-field \
-H 'Authorization: Basic XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX' \
-H 'Content-Type: application/json' \
-d '{ "key": "dc.identifier.remoteId", "value": "xxxxxxx" }'

Recupero tesi appartenenti ad una specifica collection

Il servizio restituisce l'elenco delle tesi di una collection noto il suo UUID

Endpoint: https://tesidottorato.depositolegale.it/rest/collections/<uuid_collection>/items

Method: GET

Autenticazione: Basic

Descrizione: Elenco tesi di una specifica collection

Parametri:

- limit: parametro opzionale che specifica il numero massimo di item da estrarre (se non specificato il valore di default è 100);
- offset: parametro opzionale che specifica quale debba essere il primo item da estrarre (se non specificato il valore di default è 0 cioè parte dal primo elemento);

ESEMPIO CURL:

curl 'https://test.bncf.unitesi.cineca.it/rest/collections/0738e988-cc52-409f-b08a-70c8e04f032a/items?limit=2&offset=10' \

-H 'Authorization: Basic XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX'