deploy2zenodo + deploy2inveniordm

Daniel Mohr

2026-10-08

Abstract

deploy2zenodo and deploy2inveniordm are scripts to deploy your data to Zenodo or invenio RDM providing automatic workflow via ci pipeline

Contents

deploy2zenodo + deploy2inveniordm **2**
- quick start (tldr) . . . 2
  - which script to use . . . 2
  - example CI/CD integration . . . 2
  - example bash script . . . 3
- preamble . . . 3
- intention . . . 3

deploy2zenodo **4**
- preamble . . . 4
- intention . . . 4
- how-to . . . 4
  - simple workflow . . . 4
  - very simple workflow . . . 7
  - triggered workflow . . . 7
  - complex workflow . . . 9
  - very complex workflow . . . 10
- script parameter . . . 10
  - DEPLOY2ZENODO_API_URL . . . 11
  - DEPLOY2ZENODO_ACCESS_TOKEN . . . 11
  - DEPLOY2ZENODO_DEPOSITION_ID . . . 11
  - DEPLOY2ZENODO_JSON . . . 11
  - DEPLOY2ZENODO_UPLOAD . . . 12
  - DEPLOY2ZENODO_SKIP_PUBLISH . . . 12
  - DEPLOY2ZENODO_DRYRUN . . . 13
  - DEPLOY2ZENODO_SKIPRUN . . . 13
  - DEPLOY2ZENODO_SKIP_NEW_VERSION . . . 13
  - DEPLOY2ZENODO_GET_METADATA . . . 13
  - DEPLOY2ZENODO_SKIP_UPLOAD . . . 14
  - DEPLOY2ZENODO_CURL_MAX_TIME . . . 14
  - DEPLOY2ZENODO_CURL_MAX_TIME_PUBLISH . . . 14
  - DEPLOY2ZENODO_ADD_IsCompiledBy_DEPLOY2ZENODO . . . 14
  - DEPLOY2ZENODO_ADD_IsNewVersionOf . . . 14
  - DEPLOY2ZENODO_ADD_IsPartOf . . . 15
  - DEPLOY2ZENODO_RATE_LIMIT . . . 16
- CI pipeline . . . 16
- script . . . 16

harvesting . . . 17
curating . . . 17

deploy2inveniordm is in beta **18**
preamble . . . 18
intention . . . 18
how-to . . . 18
very simple workflow (deploy2inveniordm) . . . 18
simple workflow (deploy2inveniordm) . . . 19
script parameter (deploy2inveniordm) . . . 22
DEPLOY2INVENIORDM_API_URL . . . 22
DEPLOY2INVENIORDM_ACCESS_TOKEN . . . 22
DEPLOY2INVENIORDM_DEPOSITION_ID . . . 23
DEPLOY2INVENIORDM_JSON . . . 23
DEPLOY2INVENIORDM_UPLOAD . . . 24
DEPLOY2INVENIORDM_SKIP_PUBLISH . . . 24
DEPLOY2INVENIORDM_DRYRUN . . . 25
DEPLOY2INVENIORDM_SKIPRUN . . . 25
DEPLOY2INVENIORDM_GET_METADATA . . . 25
DEPLOY2INVENIORDM_SKIP_UPLOAD . . . 25
DEPLOY2INVENIORDM_CURL_MAX_TIME . . . 25
DEPLOY2INVENIORDM_CURL_MAX_TIME_PUBLISH . . . 25
DEPLOY2INVENIORDM_ADD_IsCompiledBy_DEPLOY2INVENIORDM . . . 25
DEPLOY2INVENIORDM_RATE_LIMIT . . . 26
CI pipeline . . . 26
script . . . 27
harvesting . . . 27
curating . . . 27

license: Apache-2.0 **28**

deploy2zenodo + deploy2inveniordm

quick start (tldr)

Two lightweight shell scripts to automatically deploy your data or software to:

- Zenodo
- Invenio RDM instances

which script to use

target platform	script to use	API type

example CI/CD integration

```yaml
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/

deploy2zenodo:
  rules:
    - if: $CI_COMMIT_TAG
  variables:
    DEPLOY2ZENODO_API_URL: "https://sandbox.zenodo.org/api"
```

```
    DEPLOY2ZENODO_JSON: "metadata.json"
    DEPLOY2ZENODO_DEPOSITION_ID: "create NEW record"
    DEPLOY2ZENODO_UPLOAD: "data.zip"
  before_script:
    - apk add --no-cache curl jq
    - pubdate=$(echo "$CI_COMMIT_TIMESTAMP" | grep -Eo "^[0-9]{4}-[0-9]{2}-[0-9]{2}")
    - |
      jq -c ".metadata.version = \"$CI_COMMIT_TAG\" |
        .metadata.publication_date = \"$pubdate\"" "$DEPLOY2ZENODO_JSON" | \
        tee "tmp.json" | jq -C .
    - mv "tmp.json" "$DEPLOY2ZENODO_JSON"
```

example bash script

```
SCRIPTURL=https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/dep
curl -L "$SCRIPTURL" -o deploy2inveniordm && chmod +x deploy2inveniordm
export DEPLOY2INVENIORDM_API_URL="https://sandbox.zenodo.org/api"
 export DEPLOY2INVENIORDM_ACCESS_TOKEN="your-access-token"
export DEPLOY2INVENIORDM_JSON="metadata.json"
export DEPLOY2INVENIORDM_DEPOSITION_ID="create NEW record"
export DEPLOY2INVENIORDM_UPLOAD="data.zip"
./deploy2inveniordm
```

preamble

`deploy2zenodo` are two lightweight shell scripts to deploy your data to Zenodo or Invenio RDM, the open-source repository platform powering Zenodo. You can use them in a CI pipeline as an automatic workflow.

Environmental variables provide high flexibility. Depending on the selected flags, data can be curated

- before deployment in a merge request
- manually via the Zenodo web interface, or
- not curated at all.

Note: While `deploy2zenodo` is primarily tailored for the Zenodo API, `deploy2inveniordm` is specifically designed for the Invenio RDM API used by Zenodo.

intention

To satisfy the FAIR[1] principles[2], publications should be deposited to an open, persistent repository. By publishing to such repositories, each publication receives a PID (persistent identifier) and at least the metadata is publicly accessible, findable and citable. Furthermore, current discussions about KPIs (key performance indicator) for software and data publications also lead to the need to generate PIDs for software and data.

Especially software usually is not citable by a PID. To overcome this and make software academically significant we provide here a tool for automatic publication to the open repository zenodo.

In principal the same is true for all kind of scientific data (e. g. measurements, software and results such as papers). For every data managed in a version control system an automatic publication to an open repository is useful[3].

[1] FAIR Principles
[2] An interpretation of the FAIR principles to guide implementations in the HMC digital ecosystem.
[3] Guidance on Versioning of Digital Assets.

Software in particular is subject to frequent changes, resulting in many versions. This leads to the urge to automate the publishing process. This is not only about making the software usable through software repositories, but also about the citability of individual versions.

deploy2zenodo

preamble

deploy2zenodo is a shell script to deploy your data to Zenodo. You can use it in a CI pipeline as an automatic workflow.

Environmental variables allow very flexible use. Depending on the selected flags, the data can be curated before deployment in a merge request, in the zenodo web interface or not curated at all.

Note: deploy2zenodo is primarily designed for the Zenodo API.

intention

To satisfy the FAIR[4] principles[5], publications should be deployed to an open repository. In this way the publication gets a PID (persistent identifier) and at least the metadata is publicly accessible, findable and citable. Furthermore, current discussions about KPIs (key performance indicator) for software and data publications also lead to the need to generate PIDs for software and data.

Especially software usually is not citable by a PID. To overcome this and make software academically significant we provide here a tool for automatic publication to the open repository zenodo.

In principal the same is true for all kind of scientific data (e. g. measurements, software and results such as papers). For every data managed in a version control system an automatic publication to an open repository is useful

Software in particular is subject to frequent changes, resulting in many versions. This leads to the urge to automate the publishing process. This is not only about making the software usable through software repositories, but also about the citability of individual versions.

how-to

There are many possibilities to use deploy2zenodo but in this how-to section we will focus on a few typical use cases.

simple workflow

This workflow reflects the primary focus of deploy2zenodo.

Go to your zenodo account and create an access token.

Store it in a GitLab CI/CD variable as DEPLOY2ZENODO_ACCESS_TOKEN. Use the flags Mask variable and Protect variable. You may also choose to hide a CI/CD variable to prevent it from being revealed in the GitLab UI. Masking ensures that the variable is not displayed in the CI/CD logs, and protecting the variable limits access to authorized users and hiding reduces the risk of accidental exposure in the UI. Keep in mind the token is sensitive and private information. Therefore you should not share it or make it public available.

Then the GitLab CI/CD pipeline could look like (we use here sandbox.zenodo.org instead of zenodo.org for testing purpose):

```
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/
```

[4] FAIR Principles
[5] An interpretation of the FAIR principles to guide implementations in the HMC digital ecosystem.

```yaml
prepare_release_and_deploy2zenodo:
  stage: build
  image:
    name: alpine:latest
  variables:
    DEPLOY2ZENODO_JSON: "mymetadata.json"
  script:
    # prepare
    - TAG=$(grep version library.properties | cut -d "=" -f 2)
    - |
      echo '{"metadata":{"creators":[{"name":"family, given"}],\
        "license":{"id":"GPL-3.0-or-later"},"title":"test script alpine",\
        "version":"***","upload_type":"software"}}' | \
        jq ".metadata.version = \"$TAG\"" | tee "$DEPLOY2ZENODO_JSON"
    # prepare release
    - echo "DESCRIPTION=README.md" > variables.env
    - echo "TAG=$TAG" >> variables.env
    # prepare deploy2zenodo
    - echo "DEPLOY2ZENODO_JSON=$DEPLOY2ZENODO_JSON" >> variables.env
    - DEPLOY2ZENODO_UPLOAD="v$TAG.zip"
    - git archive --format zip --output "$DEPLOY2ZENODO_UPLOAD" "$TAG"
    - echo "DEPLOY2ZENODO_UPLOAD=$DEPLOY2ZENODO_UPLOAD" >> variables.env
  artifacts:
    reports:
      dotenv: variables.env
    paths:
      - $DEPLOY2ZENODO_JSON

release_job:
  stage: deploy
  rules:
    - if: $CI_COMMIT_TAG
      when: never
    - if: $CI_COMMIT_BRANCH == $CI_DEFAULT_BRANCH
  image:
    registry.gitlab.com/gitlab-org/release-cli:latest
  script:
    - cat /etc/os-release
  release:
    name: 'v$TAG'
    description: '$DESCRIPTION'
    tag_name: '$TAG'
    ref: '$CI_COMMIT_SHA'

deploy2zenodo:
  rules:
    - if: $CI_COMMIT_TAG
      when: never
    - if: $CI_COMMIT_BRANCH == $CI_DEFAULT_BRANCH
  variables:
    DEPLOY2ZENODO_API_URL: "https://sandbox.zenodo.org/api"
    DEPLOY2ZENODO_DEPOSITION_ID: "create NEW record"
```

We use here 3 jobs:

- The job `prepare_release_and_deploy2zenodo` prepares the variables and data for the following jobs. You can choose how to get the variables and data from your project/repository. (see hints in DEPLOY2ZENODO_JSON and DEPLOY2ZENODO_UPLOAD)
- The job `release_job` uses the workflow Create release metadata in a custom script.
- The job `deploy2zenodo` publishes the data to zenodo.

The variables are passed between the jobs using dotenv variables. And the data are passed using job artifacts.

After the first run of the above pipeline (job `deploy2zenodo`) adapt `DEPLOY2ZENODO_DEPOSITION_ID` to store the record id. Only then you are able to release new versions to zenodo.

In this example, `prepare_release_and_deploy2zenodo` always runs while the other jobs only run when the default branch is changed. This makes it possible to check the artifacts during a merge request.

The used environment variables (see script parameter) can be provided in many different ways as a GitLab CI/CD variable, e. g.:

- CI/CD variable in the UI
  - not stored in the repository
  - possible to Mask variable
  - possible to Protect variable
  - used for private data (e. g. access token)
- CI/CD variable in the .gitlab-ci.yml
  - stored in the repository
  - in public projects also publicly accessable

You should think about which information to store at which place. Here a few simple considerations:

variable	private data	note
DEPLOY2ZENODO_API_URL	no	Should a user find your publication?
DEPLOY2ZENODO_ACCESS_TOKEN	yes	Should not be shared with anyone!
DEPLOY2ZENODO_DEPOSITION_ID	no	Should a user find your publication?
DEPLOY2ZENODO_JSON	?	Is the publication public?
DEPLOY2ZENODO_UPLOAD	?	Is the publication public?

Sometimes it is easier to change the variable in the UI. For example in your first step you should set `DEPLOY2ZENODO_API_URL="https://sandbox.zenodo.org/api"` and `DEPLOY2ZENODO_DEPOSITION_ID="create NEW record"` to initiate and test your pipeline. After success you should change to `DEPLOY2ZENODO_API_URL="https://zenodo.org/api"`. And after you have created your first record, also change `DEPLOY2ZENODO_DEPOSITION_ID` to the returned value to update your dataset next time (and not create a new one). If you store these variables in the user interface, you can change them without touching your repository. On the other hand, the metadata provided via `DEPLOY2ZENODO_JSON` and the data provided via `DEPLOY2ZENODO_UPLOAD` may be created dynamically and it could therefore make sense to create these variables dynamically as well.

There are also optional variables that can help to adapt the workflow to the individual use case. For example, DEPLOY2ZENODO_SKIP_PUBLISH allows you to curate the upload to zenodo in the zenodo web interface before publishing. This is especially useful if you are setting up the workflow for the first time in your own project – but can also be used at any time.

Depending on where variables are defined, they have different priorities. For example, CI variables defined in the UI have priority and override the variables stored in the `.gitlab-ci.yml` file with

the keyword `variables`. Variables that are defined at job level, in the `script`, `before_script` or `after_script` sections, have the highest priority

An example test project is deploy2zenodo_test_simple_workflow_update.

very simple workflow

It is not necessary to create a release for publication. But we think this is the typical use case for software publication.

For a very simple workflow running when creating a tag, you could use something like:

```
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/

deploy2zenodo:
  stage: deploy
  rules:
    - if: $CI_COMMIT_TAG
  variables:
    DEPLOY2ZENODO_API_URL: "https://sandbox.zenodo.org/api"
    DEPLOY2ZENODO_JSON: "CITATION.json"
    DEPLOY2ZENODO_DEPOSITION_ID: "create NEW record"
    DEPLOY2ZENODO_UPLOAD: "$CI_PROJECT_NAME-$CI_COMMIT_TAG.zip"
    DEPLOY2ZENODO_ADD_IsCompiledBy_DEPLOY2ZENODO: "yes"
    DEPLOY2ZENODO_ADD_IsNewVersionOf: "yes"
    DEPLOY2ZENODO_ADD_IsPartOf: "yes"
    DEPLOY2ZENODO_GET_METADATA: "result.json"
  before_script:
    - env
    - apk add --no-cache cffconvert curl git jq
    - publication_date=$(echo "$CI_COMMIT_TIMESTAMP" | grep -Eo "^[0-9]{4}-[0-9]{2}-[0-9]{2}")
    - |
      cffconvert -i CITATION.cff -f zenodo | \
        jq -c '{"metadata": .} | .metadata += {"upload_type": "software"}' | \
        jq -c ".metadata.related_identifiers += [
          {\"relation\": \"isDerivedFrom\",
          \"identifier\": \"$CI_SERVER_URL/projects/$CI_PROJECT_ID\"}] |
          .metadata.version = \"$CI_COMMIT_TAG\" |
          .metadata.publication_date = \"$publication_date\"" | \
        tee "$DEPLOY2ZENODO_JSON" | jq -C .
    - git archive --format zip --output "$DEPLOY2ZENODO_UPLOAD" "$CI_COMMIT_TAG"
  artifacts:
    paths:
      - $DEPLOY2ZENODO_JSON
      - $DEPLOY2ZENODO_GET_METADATA
```

Such a simple workflow uses

- deploy_deploy2zenodo_to_zenodo in the job `deploy2zenodo` to publish itself.
- 2024-10_waw_fdm in the job `deploy2zenodo` to publish a poster.

triggered workflow

In many projects there is more than one maintainer. Therefore it is not possible to store the user token for zenodo as CI variable in the project. Otherwise, the user token would be shared with the other maintainers.

Using this triggered workflow allows to restrict the use of the user token to a specific zenodo record for other maintainers.

But the project `A` with more than one maintainer can trigger a pipeline in another (private) project `B` with only one maintainer, e. g.:

```yaml
trigger:
  stage: .post
  rules:
    - if: $CI_COMMIT_TAG
      when: never
    - if: $CI_COMMIT_BRANCH == $CI_DEFAULT_BRANCH
  image:
    name: alpine:latest
  script:
    - apk add --no-cache curl
    - curl -X POST --fail -F token="$TRIGGER_TOKEN" -F ref=main "$TRIGGER_URL"
```

Storing the `TRIGGER_TOKEN` as protected and masked CI variable (or maybe even hide CI variable) in project `A` allows any maintainer to use it and trigger the pipeline.

In the project `B` only 1 mainainer exists and you can use deploy2zenodo as normal, e. g.:

```yaml
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/

prepare_deploy2zenodo:
  image:
    name: alpine:latest
  script:
    - PROJECT_A_REPO=$(mktemp -d)
    - git clone --branch main --depth 1 "$PROJECT_A_URL" "$PROJECT_A_REPO"
    # create zip archive from latest tag
    - |
      (cd "$PROJECT_A_REPO" && \
       git archive --format zip -o "$DEPLOY2ZENODO_UPLOAD" \
       "$(git tag | sort -t "." -n -k 3 | tail -n 1)")
  artifacts:
    expire_in: 1 hrs
    paths:
      - $DEPLOY2ZENODO_UPLOAD

my_deploy2zenodo:
  extends: .deploy2zenodo
  # variables set in the script could not be overwritten by the trigger source
  before_script:
    - |
      DEPLOY2ZENODO_DEPOSITION_ID="create NEW record"
      DEPLOY2ZENODO_API_URL="https://sandbox.zenodo.org/api"
    - !reference [deploy2zenodo, before_script]

deploy2zenodo:
  rules:
    - if: '"0" == "1"'
      when: never
```

There are various ways to trigger a pipeline, e. g:

- trigger a pipeline by trigger token
- trigger using Multi-project pipelines

In the CI pipeline above the token method is used. In the CI pipeline of deploy2zenodo the multi-project pipeline is used.

Be careful: The trigger job from project `A` may overwrite variables in the triggered job from project `B`. This could lead to security concerns. Maybe Restrict who can override variables could help to overcome this.

More details: In project `A` something exists that should be published on zenodo. In project `B` the content of project `A` is published on zenodo. The pipeline in project `B` can be triggered so that this happens automatically when corresponding changes are made in project `A` (e. g. merge to default branch). Project `B` should rely as little as possible on project `A`. Unfortunately, variables can be transferred when triggering (from project `A`) and these are not trustworthy. For example, a maintainer from project `A` could pass `DEPLOY2ZENODO_API_URL` in this way and thus force communication to another server. This could cause the user token to be leaked. To avoid this, define the variable in the script – as shown in the example above. However, it is no problem to save the user token in project `B` as CI variable `DEPLOY2ZENODO_ACCESS_TOKEN`. This variable could then be overwritten from project `A`, but not read out.

Another possibility is to use Secrets management providers.

This triggered workflow is used in file_hook_server_timestamping together with deploy_file_hook_server_timestamping_to_zenodo.

complex workflow

This workflow splits the deploying to zenodo in steps. This allows to use the zenodo record (e. g. the DOI) already in the data to publish.

```yaml
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/

deploy2zenodo:
  rules:
    - if: '"0" == "1"'
      when: never

prepare_deploy2zenodo_step1:
  script:
    - ...

deploy2zenodo-step1:
  variables:
    - DEPLOY2ZENODO_SKIP_PUBLISH: "true"
    - DEPLOY2ZENODO_GET_METADATA: "newmetadata.json"
  extends: .deploy2zenodo
  image:
    name: alpine:latest
  before_script:
    - !reference [deploy2zenodo, before_script]
  script:
    - !reference [deploy2zenodo, script]
    - echo "DEPLOY2ZENODO_GET_METADATA=$DEPLOY2ZENODO_GET_METADATA" > variables.env
  artifacts:
    paths:
      - $DEPLOY2ZENODO_GET_METADATA
```

```
    reports:
      dotenv: variables.env

prepare_release:
  script:
    - echo "use the file \"$DEPLOY2ZENODO_GET_METADATA\""
    - ...

release_job:
  script:
    - ...

prepare_deploy2zenodo_step2:
  script:
    - ...

deploy2zenodo-step2:
  variables:
    DEPLOY2ZENODO_SKIP_NEW_VERSION: "true"
  extends: .deploy2zenodo
  image:
    name: alpine:latest
  before_script:
    - apk add --no-cache curl jq
```

In the step `prepare_release` you can use jq to extract data. For example the preserved DOI is available by:

```
jq .metadata.prereserve_doi.doi "$DEPLOY2ZENODO_GET_METADATA"
```

Such a complex workflow uses 2024-10_waw_fdm_talk to publish a poster. Instead of creating a release, as shown above, the poster is built in the job `build` using the DOI previously created in the job `deploy2zenodo-step1`.

very complex workflow

`deploy2zenodo` uses a combination of the triggered workflow and the complex workflow to publish itself. This is described in deploy_deploy2zenodo_to_zenodo.

script parameter

Instead of command line parameters we use environment variables.

You have to provide the following variables:

variable	content
DEPLOY2ZENODO_API_URL	The URL of the API to use.
DEPLOY2ZENODO_ACCESS_TOKEN	access token of zenodo
DEPLOY2ZENODO_DEPOSITION_ID	id of the deposition/record on zenodo
DEPLOY2ZENODO_JSON	file name with metadata in JSON format to upload
DEPLOY2ZENODO_UPLOAD	file name(s) to upload

There are other optional variables:

variable	content
DEPLOY2ZENODO_SKIP_PUBLISH	prepare record, but skip publishing
DEPLOY2ZENODO_DRYRUN	skip communicating with the external URL
DEPLOY2ZENODO_SKIPRUN	skip everything, only prints commands to execute
DEPLOY2ZENODO_SKIP_NEW_VERSION	skip creating new version
DEPLOY2ZENODO_GET_METADATA	write actual metadata to a file
DEPLOY2ZENODO_SKIP_UPLOAD	skip upload of data
DEPLOY2ZENODO_CURL_MAX_TIME	max time for curl
DEPLOY2ZENODO_CURL_MAX_TIME_PUBLISH	max time for curl during publishing
DEPLOY2ZENODO_ADD_IsCompiledBy_DEPLOY2ZENODO	reference deploy2zenodo
DEPLOY2ZENODO_ADD_IsNewVersionOf	reference previous version
DEPLOY2ZENODO_ADD_IsPartOf	reference DOI for all versions
DEPLOY2ZENODO_RATE_LIMIT	defines the wait time, default: auto

DEPLOY2ZENODO_API_URL

You can use the API of your own zenodo instance or you can use the official zenodo instance:

state	URL
production	https://zenodo.org/api
testing	https://sandbox.zenodo.org/api

DEPLOY2ZENODO_ACCESS_TOKEN

To access your zenodo account you have to provide an access token.

DEPLOY2ZENODO_DEPOSITION_ID

To update an existing record you have to provide the `id` of this record.

If you want to create a new record please set `DEPLOY2ZENODO_DEPOSITION_ID` to `create NEW record`, e. g. `DEPLOY2ZENODO_DEPOSITION_ID="create NEW record"`. After creating this record read the script output and adapt `DEPLOY2ZENODO_DEPOSITION_ID` for the next run with the returned record `id`.

DEPLOY2ZENODO_JSON

The given file should contain the metadata in JSON format.

You can write this file on your own, e. g.:

```
{
  "metadata": {
    "title": "foo",
    "upload_type": "software",
    "creators": [
      {
        "name": "ich",
        "affiliation": "bar"
      }
    ],
    "description": "foos description"
  }
}
```

You can find the necessary and possible fields on zenodo: Deposit metadata.

Or cffconvert can help harvesting the necessary metadata in JSON format from a CITATION.cff file. Unfortunately we need jq to correct the format, e. g.:

```
cffconvert -i CITATION.cff -f zenodo | \
  jq '{"metadata": .} | .metadata += {"upload_type": "software"}' | \
  tee CITATION.json
```

Since you need to adapt the output of the conversion you can also use more general tools like yq to convert a CITATION.cff file (YAML format) to JSON format.

The JSON format zenodo accepts is much more general and provides many more options than the Citation File Format. For many purposes the CITATION.cff is enough, but otherwise you can see a description of the metadata in the GitHub integration of zenodo[6] [7] [8] using `zenodo.json`, the description of the metadata in zenodo|Developers[9] or InvenioRDM[10] and the unofficial description of zenodo upload metadata schema[11].

As `description` you can use HTML. For example you could use pandoc to convert your `README.md` to HTML and jq to add the HTML code as JSON value (`jq` will escape appropriate characters if necessary):

```
pandoc -o README.html README.md
echo '{"metadata":{"title":"foo","upload_type":"software",
  "creators":[{"name":"ich","affiliation":"bar"}],
  "description":"foos description"}}' | \
  jq --rawfile README README.html '.metadata.description = $README' | \
  tee metadata.json
```

DEPLOY2ZENODO_UPLOAD

The given file(s) will be uploaded as data. Typically this would be an archive.

For example you can create an archive of a tag from a git repository:

```
TAG=0.0.3
git archive --format zip --output $TAG.zip $TAG
```

File names with spaces are not supported. Instead, if `DEPLOY2ZENODO_UPLOAD` contains space(s), the string is split at the spaces. Each individual block represents a file and these files will be uploaded.

The reason not supporting spaces is that you cannot create a CI/CD variable that is an array.

If you really not want to provide data set `DEPLOY2ZENODO_UPLOAD` to `do NOT provide data`, e. g. `DEPLOY2ZENODO_UPLOAD="do NOT provide data"`. If you want to upload 4 files with these names change the order.

Not every zenodo instance supports metadata-only records (configured by `canHaveMetadataOnlyRecords`?). For example the official zenodo instance does not allow metadata-only records! In this case an empty dummy file is uploaded. If this is the case, you should think about respecting the implicit request of the used zenodo instance to provide some data.

DEPLOY2ZENODO_SKIP_PUBLISH

If this variable is not empty the publishing step is skipped, e. g.:

[6] developers.zenodo.org GitHub
[7] github.com Referencing and citing content
[8] github: "import" past releases to Zenodo
[9] zenodo|Developers
[10] InvenioRDM: Metadata reference
[11] Zenodo upload metadata schema

```
DEPLOY2ZENODO_SKIP_PUBLISH="true"
```

Only the record is prepared – metadata and data is uploaded – but not published. You can see what will be published as a preview in the web interface of zenodo and initiate the publishing by pressing the button in the web interface.

This helps to integrate `deploy2zenodo` in your project. But you may also want to curate the upload each time before it is published.

Together with DEPLOY2ZENODO_SKIP_NEW_VERSION this allows to split deploying to zenodo in steps.

DEPLOY2ZENODO_DRYRUN

If this variable is not empty the communication to the given URL is skipped. But your parameters are analyzed. This could help to integrate `deploy2zenodo` in your project.

DEPLOY2ZENODO_SKIPRUN

If this variable is not empty nearly everything is skipped. Only the commands to be executed are echoed. This is for debugging purpose.

DEPLOY2ZENODO_SKIP_NEW_VERSION

If this variable is not empty the step creating a new version is skipped. This allows to split deploying to zenodo in steps.

Between creating a new version and deploying to zenodo you can use the zenodo record (e. g. the DOI) already in the data to publish:

```
jq .metadata.prereserve_doi.doi "$DEPLOY2ZENODO_GET_METADATA" >> README.md
```

Using a manual job allows you to first check the artifacts and data to be published before the last job run.

DEPLOY2ZENODO_GET_METADATA

If this variable is not empty the metadata of the record is stored in a file with this name. This is useful for logging or further processing after deployment.

To get these data at the end of the script an additional communication with the DEPLOY2ZENODO_API_URL server is done.

In the CI pipeline you could store the result as artifacts, e. g.:

```yaml
deploy2zenodo:
  variables:
    DEPLOY2ZENODO_GET_METADATA: "result.json"
  artifacts:
    paths:
      - $DEPLOY2ZENODO_GET_METADATA
```

You can extract values from the metadata. For example to get the DOI to site all versions:

```yaml
my_deploy2zenodo:
  extends: .deploy2zenodo
  script:
    - !reference [deploy2zenodo, script]
    - jq -r .conceptdoi "$DEPLOY2ZENODO_GET_METADATA"
```

DEPLOY2ZENODO_SKIP_UPLOAD

If this variable is not empty skip uploading the data. This is only allowed if DEPLOY2ZENODO_SKIP_PUBLISH is not empty, too.

If you split deploying to zenodo in steps using DEPLOY2ZENODO_SKIP_PUBLISH and DEPLOY2ZENODO_SKIP_NEW_VERSION you can avoid unnecessary traffic by using also DEPLOY2ZENODO_SKIP_UPLOAD.

DEPLOY2ZENODO_CURL_MAX_TIME

Max time for curl (`--max-time` flag) in seconds for normal use. Default value is 60.

DEPLOY2ZENODO_CURL_MAX_TIME_PUBLISH

Max time for curl (`--max-time` flag) in seconds during publishing. Default value is 300.

DEPLOY2ZENODO_ADD_IsCompiledBy_DEPLOY2ZENODO

If this variable is not empty a reference to deploy2zenodo is added. Something like (but with the DOI of the used version) will be added to your provided JSON file:

```
{
  "metadata": {
    "related_identifiers": [
      {
        "relation": "IsCompiledBy",
        "identifier": "10.5281/zenodo.10112959",
        "scheme": "doi",
        "resource_type": "software"
      }
    ]
  }
}
```

DEPLOY2ZENODO_ADD_IsNewVersionOf

If this variable is not empty a reference to the previous version of your record is referenced. Something like (but with the DOI of the old version and the appropriate resource_type) will be added to your provided JSON file:

```
{
  "metadata": {
    "related_identifiers": [
      {
        "relation": "IsNewVersionOf",
        "identifier": "10.5281/zenodo.10908332",
        "scheme": "doi",
        "resource_type": "software"
      }
    ]
  }
}
```

This can only work if DEPLOY2ZENODO_SKIP_NEW_VERSION is not used! If you split the run in 2 steps (the first one with DEPLOY2ZENODO_SKIP_PUBLISH and the second one with DEPLOY2ZENODO_SKIP_NEW_VERSION) you have to find the old version in the first run by yourself and provide it in the second run. This is done in deploy_deploy2zenodo_to_zenodo in the jobs

`deploy_deploy2zenodo_step1` and `deploy_deploy2zenodo_step2` to publish `deploy2zenodo`. This is something like:

```
step1:
  variables:
    DEPLOY2ZENODO_ADD_IsNewVersionOf: "yes"
  after_script:
    - |
      select(.relation==\"isNewVersionOf\") | .identifier" \
      "$DEPLOY2ZENODO_GET_METADATA")"
    - |
      select(.relation==\"isNewVersionOf\") | .resource_type" \
      "$DEPLOY2ZENODO_GET_METADATA")"
    - |
      {
      echo "LATESTDOI=$LATESTDOI"
      echo "LATESTUPLOADTYPE=$LATESTUPLOADTYPE"
      } | tee variables.env
  artifacts:
    reports:
      dotenv: variables.env

step2:
  needs:
    - job: step1
  variables:
    DEPLOY2ZENODO_SKIP_NEW_VERSION: "true"
  before_script:
    - tmpjson="$(mktemp)"
    - |
      jq ".metadata.related_identifiers += [
      {
      \"relation\":\"IsNewVersionOf\", \"identifier\":\"$LATESTDOI\",
      \"scheme\":\"doi\", \"resource_type\":\"$LATESTUPLOADTYPE\"
      }]" "$DEPLOY2ZENODO_JSON" | tee "$tmpjson"
    - mv "$tmpjson" "$DEPLOY2ZENODO_JSON"
```

Note that after_script works differently than `before_script` or `script` and does not affect the job exit code.

DEPLOY2ZENODO_ADD_IsPartOf

If this variable is not empty a reference to all versions of your record is referenced. Something like (but with the DOI of all versions and the appropriate resource_type) will be added to your provided JSON file:

```
{
  "metadata": {
    "related_identifiers": [
      {
        "relation": "IsPartOf",
        "identifier": "10.5281/zenodo.10112959",
        "scheme": "doi",
        "resource_type": "software"
      }
    ]
  }
```

```
}
```

This only works, if DEPLOY2ZENODO_DEPOSITION_ID is not given as `create NEW record`.

DEPLOY2ZENODO_RATE_LIMIT

Specifies the wait time (in seconds) between consecutive API calls.

Zenodo.org currently enforces a rate limit of maximum 30 requests per minute.

Uploading `n` files requires approximately `2n + 6` API calls:

- up to 6 calls for record creation/updating, DOI reservation, publication
- about `n` calls to delete n old files
- `n` calls to upload `n` files

Therefore, setting `DEPLOY2ZENODO_RATE_LIMIT=0` is safe for scripts that make fewer than 30 API calls.

If this variable is set to `DEPLOY2ZENODO_RATE_LIMIT=auto` (or left empty), the script estimates the total number of API calls and sets the wait time as follows:

- `0` for fewer than 23 API calls (safety margin below the limit of 30)
- `2.142857` otherwise (approx. 28 requests per minute)

CI pipeline

Using the keyword `include` it is possible to include YAML files and/or CI pipelines in your GitLab CI pipeline. In this way you can use a template of `deploy2zenodo` for your CI pipeline.

You can use the latest version deploy2zenodo.yaml in your CI pipeline. Or you can use any special versions, e. g. deploy2zenodo.yaml v0.1.0.

The provided job is called `deploy2zenodo` and you can overwrite or enhance the defined job as you need (e. g. defining when to run or defining variables).

A simple example choosing the stage to run could be:

```yaml
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/

deploy2zenodo:
  stage: deploy
```

The provided GitLab CI template of `deploy2zenodo` uses `alpine:latest` and installs necessary software curl and jq in `before_script`. To use other images you must adapt it, e. g.:

```yaml
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/

deploy2zenodo:
  image:
    name: almalinux:latest
  before_script:
    - echo "nothing to do"
```

script

You can use the script directly. But that is not our main focus of `deploy2zenodo`, so we keep it short. For example:

```
SCRIPTURL=https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/dep
export DEPLOY2ZENODO_API_URL=https://sandbox.zenodo.org/api
export DEPLOY2ZENODO_ACCESS_TOKEN=***
export DEPLOY2ZENODO_DEPOSITION_ID="create NEW record"
export DEPLOY2ZENODO_JSON=metadata.json
export DEPLOY2ZENODO_UPLOAD="foo.zip bar.md"
export DEPLOY2ZENODO_SKIP_PUBLISH="true"
export DEPLOY2ZENODO_DRYRUN=""
export DEPLOY2ZENODO_SKIPRUN=""
export DEPLOY2ZENODO_SKIP_NEW_VERSION=""
export DEPLOY2ZENODO_GET_METADATA="upload.json"
export DEPLOY2ZENODO_SKIP_UPLOAD=""
export DEPLOY2ZENODO_CURL_MAX_TIME=""
export DEPLOY2ZENODO_CURL_MAX_TIME_PUBLISH=""
export DEPLOY2ZENODO_ADD_IsCompiledBy_DEPLOY2ZENODO="yes"
export DEPLOY2ZENODO_ADD_IsNewVersionOf=""
export DEPLOY2ZENODO_ADD_IsPartOf="yes"
curl -L "$SCRIPTURL" | tee deploy2zenodo.sh | sh
```

It's worth noting that we try to handle private information such as the access token in a secure way in the script, it is still a sensitive piece of information that should not be shared with anyone who does not need access to the corresponding Zenodo account.

Important: Using the script as described in this section (e. g. on a desktop computer) does not allow for the masking of CI variables, which can expose sensitive information such as access tokens. We recommend that users take steps to mitigate this risk.

harvesting

As already mentioned you have to provide the metadata and the data to upload.

In my opinion, this is very dependent on the project. In many programming languages, there is a convention to store metadata such as name, author, description, version and license in certain files (`pyproject.toml`, `library.properties`, ...). In order to deploy this information to zenodo, it must be available in a certain format and with a certain vocabulary. The already mentioned cffconvert tries to do this at least for cff files. Other tools such as somesy have a somewhat different focus, but they can also help in a pipeline/toolchain. For example, you could use it to convert a typical python `pyproject.toml` into `CITATION.cff` and then use cffconvert and jq to get metadata for zenodo. However, hermes should also be mentioned here. hermes tries to merge metadata from different sources and to provide it for zenodo.

curating

You have various options for curating the data for publication.

The typical workflow in software development is to work in developer or feature branches and then merge them with the default branch (e. g. `main`). This is usually done in a merge request. If the harvesting of metadata and data is already taking place in a CI pipeline at this point, this can also be checked in the merge request.

If publishing is prevented by using `DEPLOY2ZENODO_SKIP_PUBLISH`, the preview in the zenodo web interface can be used to check the result.

If you have implemented a stable, functioning process, curation can also be omitted and publishing can be fully automated.

deploy2inveniordm is in beta

preamble

deploy2inveniordm is a shell script to deploy your data to Zenodo using the Invenio RDM API. You can use it in a CI pipeline as an automatic workflow.

Environmental variables allow very flexible use. Depending on the selected flags, the data can be curated before deployment in a merge request, in the zenodo web interface or not curated at all.

Note: deploy2inveniordm is designed for the Invenio RDM API used by Zenodo. It may also work with Invenio RDM, but compatibility is not guaranteed or tested. Test the script with any new API before using it.

intention

To satisfy the FAIR[12] principles[13], publications should be deployed to an open repository. In this way the publication gets a PID (persistent identifier) and at least the metadata is publicly accessible, findable and citable. Furthermore, current discussions about KPIs (key performance indicator) for software and data publications also lead to the need to generate PIDs for software and data.

Especially software usually is not citable by a PID. To overcome this and make software academically significant we provide here a tool for automatic publication to the open repository zenodo.

In principal the same is true for all kind of scientific data (e. g. measurements, software and results such as papers). For every data managed in a version control system an automatic publication to an open repository is useful

Software in particular is subject to frequent changes, resulting in many versions. This leads to the urge to automate the publishing process. This is not only about making the software usable through software repositories, but also about the citability of individual versions.

how-to

There are many ways to use deploy2inveniordm. Below, we focus on two typical use cases: a very simple workflow (deploy2inveniordm) and a more structured, production-ready workflow simple workflow (deploy2inveniordm).

very simple workflow (deploy2inveniordm)

Maybe this very simple workflow to simplified. But it is a good starting point for automating deployments to Zenodo via GitLab CI/CD pipeline.

Go to your zenodo account and create an access token.

Store it as a GitLab CI/CD variable named DEPLOY2INVENIORDM_ACCESS_TOKEN. Use the flag Mask variable to ensure the variable value is not displayed in CI/CD job logs. Additionally, enable the Protect variable option to restrict access to authorized users only. You may also choose to hide a CI/CD variable to prevent it from being revealed in the GitLab UI. Masking prevents the variable from appearing in logs, while protection limits access to trusted users. Hiding reduces the risk of accidental exposure in the UI. Since this token contains sensitive and private information, it must never be shared or made publicly available.

Then the very simple workflow running when creating a tag, could look like (we use here sandbox.zenodo.org instead of zenodo.org for testing purpose):

```
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/
```

[12] FAIR Principles
[13] An interpretation of the FAIR principles to guide implementations in the HMC digital ecosystem.

```yaml
deploy2inveniordm:
  stage: deploy
  rules:
    - if: $CI_COMMIT_TAG
  variables:
    DEPLOY2INVENIORDM_API_URL: "https://sandbox.zenodo.org/api"
    DEPLOY2INVENIORDM_JSON: "metadata.json"
    DEPLOY2INVENIORDM_DEPOSITION_ID: "create NEW record"
    DEPLOY2INVENIORDM_UPLOAD: "$CI_PROJECT_NAME-$CI_COMMIT_TAG.zip"
    DEPLOY2INVENIORDM_ADD_IsCompiledBy_DEPLOY2INVENIORDM: "yes"
    DEPLOY2INVENIORDM_GET_METADATA: "result.json"
  before_script:
    - env
    - apk add --no-cache curl git jq
    - publication_date=$(echo "$CI_COMMIT_TIMESTAMP" | grep -Eo "^[0-9]{4}-[0-9]{2}-[0-9]{2}")
    - |
        echo '{"metadata":{"creators":
          [{"person_or_org":{"type":"personal",
          "given_name":"given","family_name":"family"},
          "affiliations":[{"id":"04bwf3e34"}]}],
          "rights":[{"id":"cc-by-4.0","props":{"scheme":"spdx"}}],
          "title":"test script","version":"***",
          "resource_type":{"id":"software"},"publisher":"Zenodo"}}' | \
          jq -c ".metadata.related_identifiers += [
          {\"relation\": \"isDerivedFrom\",
          \"identifier\": \"$CI_SERVER_URL/projects/$CI_PROJECT_ID\"}] |
          .metadata.version = \"$CI_COMMIT_TAG\" |
          .metadata.publication_date = \"$publication_date\"" | \
          tee "$DEPLOY2INVENIORDM_JSON" | jq -C .
    - git archive --format zip --output "$DEPLOY2INVENIORDM_UPLOAD" "$CI_COMMIT_TAG"
  artifacts:
    paths:
      - $DEPLOY2INVENIORDM_JSON
      - $DEPLOY2INVENIORDM_GET_METADATA
```

Such a simple workflow uses

- 2026_hmc_poster_deploy2zenodo_deploy2inveniordm

simple workflow (deploy2inveniordm)

This workflow reflects the primary focus of `deploy2inveniordm`.

This workflow separates preparation from release and deployment, enabling better control and auditability.

The GitLab CI/CD pipeline could look like:

```yaml
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/

prepare_release_and_deploy2inveniordm:
  stage: build
  image:
    name: alpine:latest
  variables:
```

```
    DEPLOY2INVENIORDM_JSON: "mymetadata.json"
  script:
    # prepare
    - TAG=$(grep version library.properties | cut -d "=" -f 2)
    - |
        echo '{"metadata":{"creators":
          [{"person_or_org":{"type":"personal",
          "given_name":"given","family_name":"family"},
          "affiliations":[{"id":"04bwf3e34"}]}],
          "rights":[{"id":"cc-by-4.0","props":{"scheme":"spdx"}}],
          "title":"test script","version":"***",
          "resource_type":{"id":"software"},"publisher":"Zenodo"}}' | \
          jq ".metadata.version = \"$TAG\"" | tee "$DEPLOY2INVENIORDM_JSON"
    # prepare release
    - echo "DESCRIPTION=README.md" > variables.env
    - echo "TAG=$TAG" >> variables.env
    # prepare deploy2inveniordm
    - echo "DEPLOY2INVENIORDM_JSON=$DEPLOY2INVENIORDM_JSON" >> variables.env
    - DEPLOY2INVENIORDM_UPLOAD="v$TAG.zip"
    - git archive --format zip --output "$DEPLOY2INVENIORDM_UPLOAD" "$TAG"
    - echo "DEPLOY2INVENIORDM_UPLOAD=$DEPLOY2INVENIORDM_UPLOAD" >> variables.env
  artifacts:
    reports:
      dotenv: variables.env
    paths:
      - $DEPLOY2INVENIORDM_JSON

release_job:
  stage: deploy
  rules:
    - if: $CI_COMMIT_TAG
      when: never
    - if: $CI_COMMIT_BRANCH == $CI_DEFAULT_BRANCH
  image:
    registry.gitlab.com/gitlab-org/release-cli:latest
  script:
    - cat /etc/os-release
  release:
    name: 'v$TAG'
    description: '$DESCRIPTION'
    tag_name: '$TAG'
    ref: '$CI_COMMIT_SHA'

deploy2inveniordm:
  rules:
    - if: $CI_COMMIT_TAG
      when: never
    - if: $CI_COMMIT_BRANCH == $CI_DEFAULT_BRANCH
  variables:
    DEPLOY2INVENIORDM_API_URL: "https://sandbox.zenodo.org/api"
    DEPLOY2INVENIORDM_DEPOSITION_ID: "create NEW record"
```

We use here 3 jobs:

- The job `prepare_release_and_deploy2inveniordm` prepares the variables and data for the following jobs. You can choose how to get the variables and data from your project/repository.

- The job `release_job` uses the workflow Create release metadata in a custom script.
- The job `deploy2inveniordm` publishes the data to zenodo.

The variables are passed between the jobs using dotenv variables. And the data are passed using job artifacts.

After the first run of the above pipeline (job `deploy2inveniordm`) adapt `DEPLOY2INVENIORDM_DEPOSITION_ID` to store the record id. Only then you are able to release new versions to zenodo.

In this example, `prepare_release_and_deploy2inveniordm` always runs while the other jobs only run when the default branch is changed. This makes it possible to check the artifacts during a merge request.

The used environment variables can be provided in many different ways as a GitLab CI/CD variable, e. g.:

- CI/CD variable in the UI
  - not stored in the repository
  - possible to Mask variable
  - possible to Protect variable
  - possible to Hide variable
  - used for private data (e. g. access token)
- CI/CD variable in the .gitlab-ci.yml
  - stored in the repository
  - in public projects also publicly accessable

You should think about which information to store at which place. Here a few simple considerations:

variable	private data	note
DEPLOY2INVENIORDM_API_URL	no	Should a user find your publication?
DEPLOY2INVENIORDM_ACCESS_TOKEN	yes	Should not be shared with anyone!
DEPLOY2INVENIORDM_DEPOSITION_ID	no	Should a user find your publication?
DEPLOY2INVENIORDM_JSON	?	Is the publication public?
DEPLOY2INVENIORDM_UPLOAD	?	Is the publication public?

Store sensitive data (like tokens) in the GitLab UI with masking, protection, and hiding enabled.

Sometimes it is easier to change the variable in the UI. For example in your first step you should set `DEPLOY2INVENIORDM_API_URL="https://sandbox.zenodo.org/api"` and `DEPLOY2INVENIORDM_DEPOSITION_ID="create NEW record"` to initiate and test your pipeline. After success you should change to `DEPLOY2INVENIORDM_API_URL="https://zenodo.org/api"`. And after you have created your first record, also change `DEPLOY2INVENIORDM_DEPOSITION_ID` to the returned value to update your dataset next time (and not create a new one). If you store these variables in the user interface, you can change them without touching your repository. On the other hand, the metadata provided via `DEPLOY2INVENIORDM_JSON` and the data provided via `DEPLOY2INVENIORDM_UPLOAD` may be created dynamically and it could therefore make sense to create these variables dynamically as well.

There are also optional variables that can help to adapt the workflow to the individual use case. For example, `DEPLOY2INVENIORDM_SKIP_PUBLISH` allows you to curate the upload to zenodo in the zenodo web interface before publishing. This is especially useful if you are setting up the workflow for the first time in your own project – but can also be used at any time.

Optional Variables:

- `DEPLOY2INVENIORDM_SKIP_PUBLISH`: useful for manual curation in the Zenodo web interface
- `DEPLOY2INVENIORDM_DRYRUN`: useful for testing and debugging a pipeline
- `DEPLOY2INVENIORDM_SKIPRUN`: useful for debugging and verifying the workflow logic
- `DEPLOY2INVENIORDM_GET_METADATA`: useful for extrating DOI for use in later steps
- ...

Depending on where variables are defined, they have different priorities. For example, CI variables defined in the UI have priority and override the variables stored in the `.gitlab-ci.yml` file with the keyword `variables`. Variables that are defined at job level, in the `script`, `before_script` or `after_script` sections, have the highest priority

script parameter (deploy2inveniordm)

Instead of command line parameters we use environment variables.

You have to provide the following variables:

variable	content
DEPLOY2INVENIORDM_API_URL	The URL of the API to use.
DEPLOY2INVENIORDM_ACCESS_TOKEN	access token of zenodo
DEPLOY2INVENIORDM_DEPOSITION_ID	id of the deposition/record on zenodo
DEPLOY2INVENIORDM_JSON	file name with metadata in JSON format to upload
DEPLOY2INVENIORDM_UPLOAD	file name(s) to upload

There are other optional variables:

variable	content
DEPLOY2INVENIORDM_SKIP_PUBLISH	prepare record, but skip publishing
DEPLOY2INVENIORDM_DRYRUN	skip communicating with the external URL
DEPLOY2INVENIORDM_SKIPRUN	skip everything, only prints commands to execute
DEPLOY2INVENIORDM_GET_METADATA	write actual metadata to a file
DEPLOY2INVENIORDM_SKIP_UPLOAD	skip upload of data
DEPLOY2INVENIORDM_CURL_MAX_TIME	max time for curl
DEPLOY2INVENIORDM_CURL_MAX_TIME_PUBLISH	max time for curl during publishing
DEPLOY2INVENIORDM_ADD_IsCompiledBy_reference	DEPLOY2INVENIORDM deploy2zenodo
DEPLOY2INVENIORDM_RATE_LIMIT	defines the wait time, default: auto

DEPLOY2INVENIORDM_API_URL

You can use the API of your own zenodo instance or you can use the official zenodo instance:

state	URL
production	https://zenodo.org/api
testing	https://sandbox.zenodo.org/api

DEPLOY2INVENIORDM_ACCESS_TOKEN

To access your zenodo account you have to provide an access token.

DEPLOY2INVENIORDM_DEPOSITION_ID

To update an existing record you have to provide the `id` of this record.

If you want to create a new record please set `DEPLOY2INVENIORDM_DEPOSITION_ID` to `create NEW record`, e. g. `DEPLOY2INVENIORDM_DEPOSITION_ID="create NEW record"`. After creating this record read the script output and adapt `DEPLOY2INVENIORDM_DEPOSITION_ID` for the next run with the returned record `id`.

DEPLOY2INVENIORDM_JSON

The given file should contain the metadata in JSON format.

You can write this file on your own, e. g.:

```
{
  "metadata": {
    "creators": [
      {
        "person_or_org": {
          "type": "personal",
          "given_name": "given",
          "family_name": "family"
        },
        "affiliations": [
          {
            "id": "04bwf3e34",
            "name": "Deutsches Zentrum für Luft- und Raumfahrt e. V."
          }
        ]
      }
    ],
    "rights": [
      {
        "id": "cc-by-4.0",
        "props": {
          "scheme": "spdx"
        }
      }
    ],
    "title": "foo",
    "version": "***",
    "resource_type": {
      "id": "software"
    },
    "publisher": "Zenodo",
    "description": "foos description"
  }
}
```

You can find the necessary and possible fields on inveniordm: minimal set of metadata.

Or cffconvert can help harvesting the necessary metadata in JSON format from a CITATION.cff file. Unfortunately we need jq to correct the format, e. g. (there are more corrections necessary):

```
cffconvert -i CITATION.cff -f zenodo | \
  jq '{"metadata": .} | .metadata += {"upload_type": "software"}' | \
  tee CITATION.json
```

Since you need to adapt the output of the conversion you can also use more general tools like yq to convert a CITATION.cff file (YAML format) to JSON format.

The JSON format zenodo accepts is much more general and provides many more options than the Citation File Format. For many purposes the CITATION.cff is enough, but otherwise you can see a description of the metadata in the GitHub integration of zenodo[14] [15] [16] using `zenodo.json`, the description of the metadata in zenodo|Developers[17] or InvenioRDM[18] and the unofficial description of zenodo upload metadata schema[19].

As `description` you can use HTML. For example you could use pandoc to convert your `README.md` to HTML and jq to add the HTML code as JSON value (`jq` will escape appropriate characters if necessary):

```
pandoc -o README.html README.md
echo '{"metadata":{"title":"foo","upload_type":"software",
  "creators":[{"name":"ich","affiliation":"bar"}],
  "description":"foos description"}}' | \
  jq --rawfile README README.html '.metadata.description = $README' | \
  tee metadata.json
```

DEPLOY2INVENIORDM_UPLOAD

The given file(s) will be uploaded as data. Typically this would be an archive.

For example you can create an archive of a tag from a git repository:

```
TAG=0.0.3
git archive --format zip --output $TAG.zip $TAG
```

File names with spaces are not supported. Instead, if `DEPLOY2INVENIORDM_UPLOAD` contains space(s), the string is split at the spaces. Each individual block represents a file and these files will be uploaded.

The reason not supporting spaces is that you cannot create a CI/CD variable that is an array.

If you really not want to provide data set `DEPLOY2INVENIORDM_UPLOAD` to `do NOT provide data`, e. g. `DEPLOY2INVENIORDM_UPLOAD="do NOT provide data"`. If you want to upload 4 files with these names change the order.

Not every zenodo instance supports metadata-only records (configured by `canHaveMetadataOnlyRecords`?). For example the official zenodo instance does not allow metadata-only records! In this case an empty dummy file is uploaded. If this is the case, you should think about respecting the implicit request of the used zenodo instance to provide some data.

DEPLOY2INVENIORDM_SKIP_PUBLISH

If this variable is not empty the publishing step is skipped, e. g.:

```
DEPLOY2INVENIORDM_SKIP_PUBLISH="true"
```

Only the record is prepared – metadata and data is uploaded – but not published. You can see what will be published as a preview in the web interface of zenodo and initiate the publishing by pressing the button in the web interface.

This helps to integrate `deploy2inveniordm` in your project. But you may also want to curate the upload each time before it is published.

[14] developers.zenodo.org GitHub
[15] github.com Referencing and citing content
[16] github: "import" past releases to Zenodo
[17] zenodo|Developers
[18] InvenioRDM: Metadata reference
[19] Zenodo upload metadata schema

This allows to split deploying to zenodo in steps.

DEPLOY2INVENIORDM_DRYRUN

If this variable is not empty the communication to the given URL is skipped. But your parameters are analyzed. This could help to integrate `deploy2inveniordm` in your project.

DEPLOY2INVENIORDM_SKIPRUN

If this variable is not empty nearly everything is skipped. Only the commands to be executed are echoed. This is for debugging purpose.

DEPLOY2INVENIORDM_GET_METADATA

If this variable is not empty the metadata of the record is stored in a file with this name. This is useful for logging or further processing after deployment.

To get these data at the end of the script an additional communication with the DEPLOY2INVENIORDM_API_URL server is done.

In the CI pipeline you could store the result as artifacts, e. g.:

```
deploy2inveniordm:
  variables:
    DEPLOY2INVENIORDM_GET_METADATA: "result.json"
  artifacts:
    paths:
      - $DEPLOY2INVENIORDM_GET_METADATA
```

You can extract values from the metadata. For example to get the DOI to site all versions:

```
my_deploy2inveniordm:
  extends: .deploy2inveniordm
  script:
    - !reference [deploy2inveniordm, script]
    - jq -r .conceptdoi "$DEPLOY2INVENIORDM_GET_METADATA"
```

DEPLOY2INVENIORDM_SKIP_UPLOAD

If this variable is not empty skip uploading the data. This is only allowed if DEPLOY2INVENIORDM_SKIP_PUBLISH is not empty, too.

If you split deploying to zenodo in steps using DEPLOY2INVENIORDM_SKIP_PUBLISH you can avoid unnecessary traffic by using also DEPLOY2INVENIORDM_SKIP_UPLOAD.

DEPLOY2INVENIORDM_CURL_MAX_TIME

Max time for curl (`--max-time` flag) in seconds for normal use. Default value is 60.

DEPLOY2INVENIORDM_CURL_MAX_TIME_PUBLISH

Max time for curl (`--max-time` flag) in seconds during publishing. Default value is 300.

DEPLOY2INVENIORDM_ADD_IsCompiledBy_DEPLOY2INVENIORDM

If this variable is not empty a reference to deploy2inveniordm is added. Something like (but with the DOI of the used version) will be added to your provided JSON file:

```
{
  "metadata": {
    "related_identifiers": [
      {
        "relation": "IsCompiledBy",
        "identifier": "10.5281/zenodo.10112959",
        "scheme": "doi",
        "resource_type": "software"
      }
    ]
  }
}
```

DEPLOY2INVENIORDM_RATE_LIMIT

Specifies the wait time (in seconds) between consecutive API calls.

Zenodo.org currently enforces a rate limit of maximum 30 requests per minute.

Uploading `n` files requires approximately `4n + 6` API calls:

- up to 6 calls for record creation/updating, DOI reservation, publication
- about `n` calls to delete n old files
- `3n` calls to upload `n` files

Therefore, setting `DEPLOY2INVENIORDM_RATE_LIMIT=0` is safe for scripts that make fewer than 30 API calls.

If this variable is set to `DEPLOY2INVENIORDM_RATE_LIMIT=auto` (or left empty), the script estimates the total number of API calls and sets the wait time as follows:

- `0` for fewer than 23 API calls (safety margin below the limit of 30)
- `2.142857` otherwise (approx. 28 requests per minute)

If you are using a different instance without a rate limit, you can set `DEPLOY2INVENIORDM_RATE_LIMIT=0` to skip the wait time.

CI pipeline

Using the keyword `include` it is possible to include YAML files and/or CI pipelines in your GitLab CI pipeline. In this way you can use a template of `deploy2inveniordm` for your CI pipeline.

You can use the latest version deploy2inveniordm.yaml in your CI pipeline. Or you can use any special versions, e. g. deploy2inveniordm.yaml v0.1.0.

The provided job is called `deploy2inveniordm` and you can overwrite or enhance the defined job as you need (e. g. defining when to run or defining variables).

A simple example choosing the stage to run could be:

```
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/

deploy2inveniordm:
  stage: deploy
```

The provided GitLab CI template of `deploy2inveniordm` uses `alpine:latest` and installs necessary software curl and jq in `before_script`. To use other images you must adapt it, e. g.:

```
include:
  - remote: 'https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/
```

```yaml
deploy2inveniordm:
  image:
    name: almalinux:latest
  before_script:
    - echo "nothing to do"
```

script

You can use the script directly. But that is not our main focus of `deploy2inveniordm`, so we keep it short. For example:

```
SCRIPTURL="https://gitlab.com/deploy2zenodo/deploy2zenodo/-/releases/permalink/latest/downloads/de
export DEPLOY2INVENIORDM_API_URL="https://sandbox.zenodo.org/api"
export DEPLOY2INVENIORDM_ACCESS_TOKEN=***
export DEPLOY2INVENIORDM_DEPOSITION_ID="create NEW record"
export DEPLOY2INVENIORDM_JSON="metadata.json"
export DEPLOY2INVENIORDM_UPLOAD="foo.zip bar.md"
export DEPLOY2INVENIORDM_SKIP_PUBLISH="true"
export DEPLOY2INVENIORDM_DRYRUN=""
export DEPLOY2INVENIORDM_SKIPRUN=""
export DEPLOY2INVENIORDM_GET_METADATA="upload.json"
export DEPLOY2INVENIORDM_SKIP_UPLOAD=""
export DEPLOY2INVENIORDM_CURL_MAX_TIME=""
export DEPLOY2INVENIORDM_CURL_MAX_TIME_PUBLISH=""
export DEPLOY2INVENIORDM_ADD_IsCompiledBy_DEPLOY2INVENIORDM="yes"
curl -L "$SCRIPTURL" | tee deploy2inveniordm.sh | sh
```

It's worth noting that we try to handle private information such as the access token in a secure way in the script, it is still a sensitive piece of information that should not be shared with anyone who does not need access to the corresponding Zenodo account.

Important: Using the script as described in this section (e. g. on a desktop computer) does not allow for the masking of CI variables, which can expose sensitive information such as access tokens. We recommend that users take steps to mitigate this risk.

harvesting

As already mentioned you have to provide the metadata and the data to upload.

In my opinion, this is very dependent on the project. In many programming languages, there is a convention to store metadata such as name, author, description, version and license in certain files (`pyproject.toml`, `library.properties`, ...). In order to deploy this information to zenodo, it must be available in a certain format and with a certain vocabulary. The already mentioned cffconvert tries to do this at least for cff files. Other tools such as somesy have a somewhat different focus, but they can also help in a pipeline/toolchain. For example, you could use it to convert a typical python `pyproject.toml` into `CITATION.cff` and then use cffconvert and jq to get metadata for zenodo. However, hermes should also be mentioned here. hermes tries to merge metadata from different sources and to provide it for zenodo.

curating

You have various options for curating the data for publication.

The typical workflow in software development is to work in developer or feature branches and then merge them with the default branch (e. g. `main`). This is usually done in a merge request. If the harvesting of metadata and data is already taking place in a CI pipeline at this point, this can also be checked in the merge request.

If publishing is prevented by using `DEPLOY2INVENIORDM_SKIP_PUBLISH`, the preview in the zenodo web interface can be used to check the result.

If you have implemented a stable, functioning process, curation can also be omitted and publishing can be fully automated.

license: Apache-2.0

`deploy2zenodo` and `deploy2inveniordm` are licensed under the Apache-2.0.


```
Copyright 2023-2026 Daniel Mohr and
   Deutsches Zentrum fuer Luft- und Raumfahrt e. V., D-51170 Koeln

Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.
```