

# Proces for deling af data i GBIF med eget data repository og autosync

1. I skal lave et data dump fra jeres database, hvor dataformatet henholder sig til et sæt regler for bl.a. dato, videnskabeligt navn, længde- og breddegrader i decimal-grader osv. Tillige skal der være nogle obligatoriske kolonner til stede (unikt ID pr. datapost osv.), som alt sammen er beskrevet under den såkaldte Darwin Core Standard.
  - a. Hvad er et Darwin Core arkiv:
    - i. [https://gcube.wiki.gcube-system.org/gcube/Darwin\\_Core\\_Terms](https://gcube.wiki.gcube-system.org/gcube/Darwin_Core_Terms)
    - ii. <https://ipt.gbif.org/manual/en/ipt/2.5/dwca-guide#what-is-darwin-core-archive-dwc-a>

Der skal altså genereres en eller flere tsv/csv filer med data - i mange tilfælde bare to: én til dataposter med fund/indsamling og én med link til billeder, deres ophavsperson, brugslicens osv.

Ud over disse filer skal man have:

2. en meta.xml der beskriver kolonnerne i datafilerne (rækkefølgen på kolonnehoveder for data navngivet med de kolonnenavne, der anvendes i Darwin Core formatet)
3. en eml.xml som er en "about" fil til datasættet (struktureret indhold der beskriver datasættet i ord, altså fx lidt om organisationen/foreningen, hvem der er kontaktperson(er), website URL, link til logo osv.)

Se også <https://ipt.gbif.org/manual/en/ipt/2.5/dwca-guide#dwc-a-components>

Anbefalede (minimum) kolonner for occurrence data:

<https://ipt.gbif.org/manual/en/ipt/2.5/occurrence-data#templates>

Hvis man vil slå specifikke Darwin Core termer op, henvises til <https://dwc.tdwg.org/terms/>. Dette er den fulde liste af Darwin Core termer - og her hjælper DanBIF med at udpege, hvilke I har brug for, der er næsten ingen grænser for, hvad der kan passes ind i Darwin Core template.

4. Når man har lavet et script til data dump og komprimering af datafiler+meta.xml+eml.xml, placeres zip arkivet på en server, hvor det er tilgængeligt via http(s)

Så kan GBIF, med DanBIFs hjælp, lave en test crawl og verificere, at det ser rigtigt ud.

Man kan inden da bruge data validatoren til at verificere om et Darwin Core arkiv er korrekt struktureret: <https://www.gbif.org/tools/data-validator>

Eksempler på datasæt med eget data repository og autosync i GBIF:

Arter.dk: <https://www.gbif.org/dataset/963a6b96-4d22-4428-86e4-afee52cf4a8e>

DOF: <https://www.gbif.org/dataset/95db4db8-f762-11e1-a439-00145eb45e9a>

Svampeforening: <https://www.gbif.org/dataset/84d26682-f762-11e1-a439-00145eb45e9a>