# FAIR DIGITAL OBJECTS WITH STRUCTURED TRAIT DATA IN THE GFBIO PORTAL

Anton Link[1], Stefan Seifert[1], Tanja Weibulat[1,2],
Wolfgang Reichert[1], Markus Weiss[1], Dagmar Triebel[1]
[1]SNSB IT Center, [2]GFBio e.V.

## EML METADATA AND STRUCTURED DATA

The ecological and environmental research community relies on the **Ecological Markup Language** (EML) as metadata exchange standard. It allows for providing information packages with XML-based EML files and assigned **tabular CSV data files** structured according user-specific proprietary content schemas, often matrices with measurement data (EML-CSV).

The biological systematics, nature observation and conservation domain focus on structuring descriptive and trait data acording community agreed vocabularies, internal data schemas and classifications. A combination of EML metadata and research data in the XML-based TDWG approved standard **SDD** (Structured Descriptive Data) is appropriate to create well **structured trait data** (EML-SDD).

## DATA MANAGEMENT IN DWB

The data management in **Diversity Workbench** (DWB) software combines the EML metadata automatically provided by the **GFBio submission tool** with manual data curation of structured descriptive data from various domains and disciplines .

DiversityDescriptions (DD), one of the DWB modules, is able to generate export files in DELTA, SDD and CSV format (fig. 2). There are DWB cross-module export functions reading metadata stored in DiversityProjects and combining them with exports from DD into **EML-CSV respectively EML-SDD files** packed in zip archives.
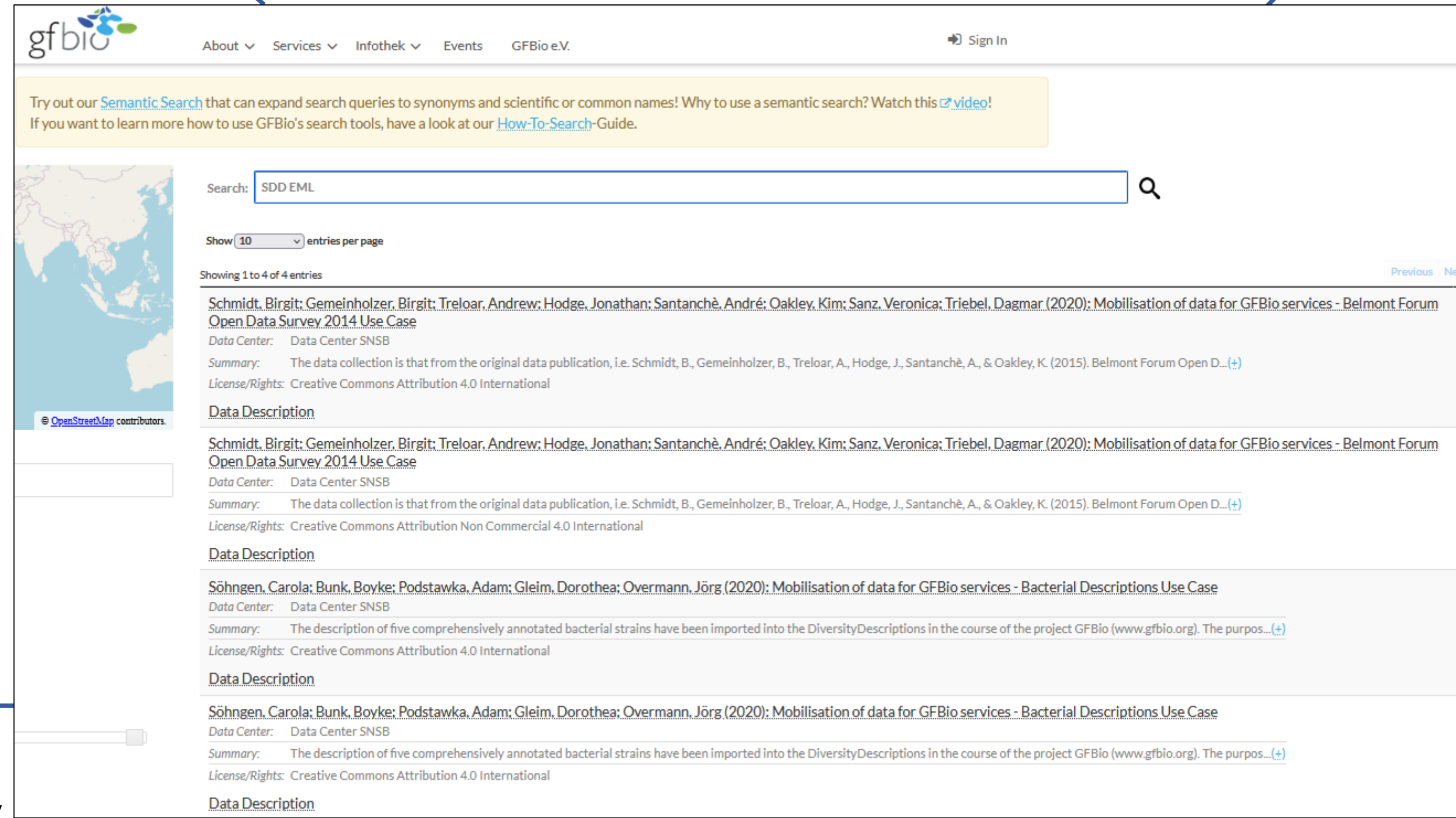


*Figure 2: DWB-DD export interface*



*Figure 1: Access to EML-SDD zip archives (use cases) via the GFBio Data Portal*

## TOWARDS FAIR DIGITAL OBJECTS

The work on creating FAIR EML-SDD structured digital objects published with DOI **will be continued as part of NFDI4Biodiversity**. The packages are ready to deliver structured (trait) data reusable through ecological and biological analysis and identification software and platforms like Morpho, knb DataONE, xper3, DELTA and, in near future, DiversityNaviKey; see also Harjes et al 2020 (fig. 3).



*Figure 3: Harjes et al. (2020), DOI: 10.1093/database/baaa059*

## TOWARDS FAIR DATA PUBLICATION

The **DataCite FABRICA** web interface (fig. 4) with DOI assignment and OAI-PMH API is used for data publication. DataCite Dublin Core metadata (+ EML elements) guide to persistent access of zip archives with EML-CSV files and EML-SDD files and URI of the landing pages at the SNSB. The **GFBio data portal** is addressing the same service (fig. 1, use cases).



*Figure 4: DataCite Fabrica interface (https://doi.datacite.org/ )*

## TOWARDS SNSB LANDING PAGES

The dynamic generation of **SNSB landing pages** with primary access to generated zip archives and their **digital object identifiers (DOIs),** as well as the linking of versioning and archival information is realised in a prototypic version with use cases (fig. 5).
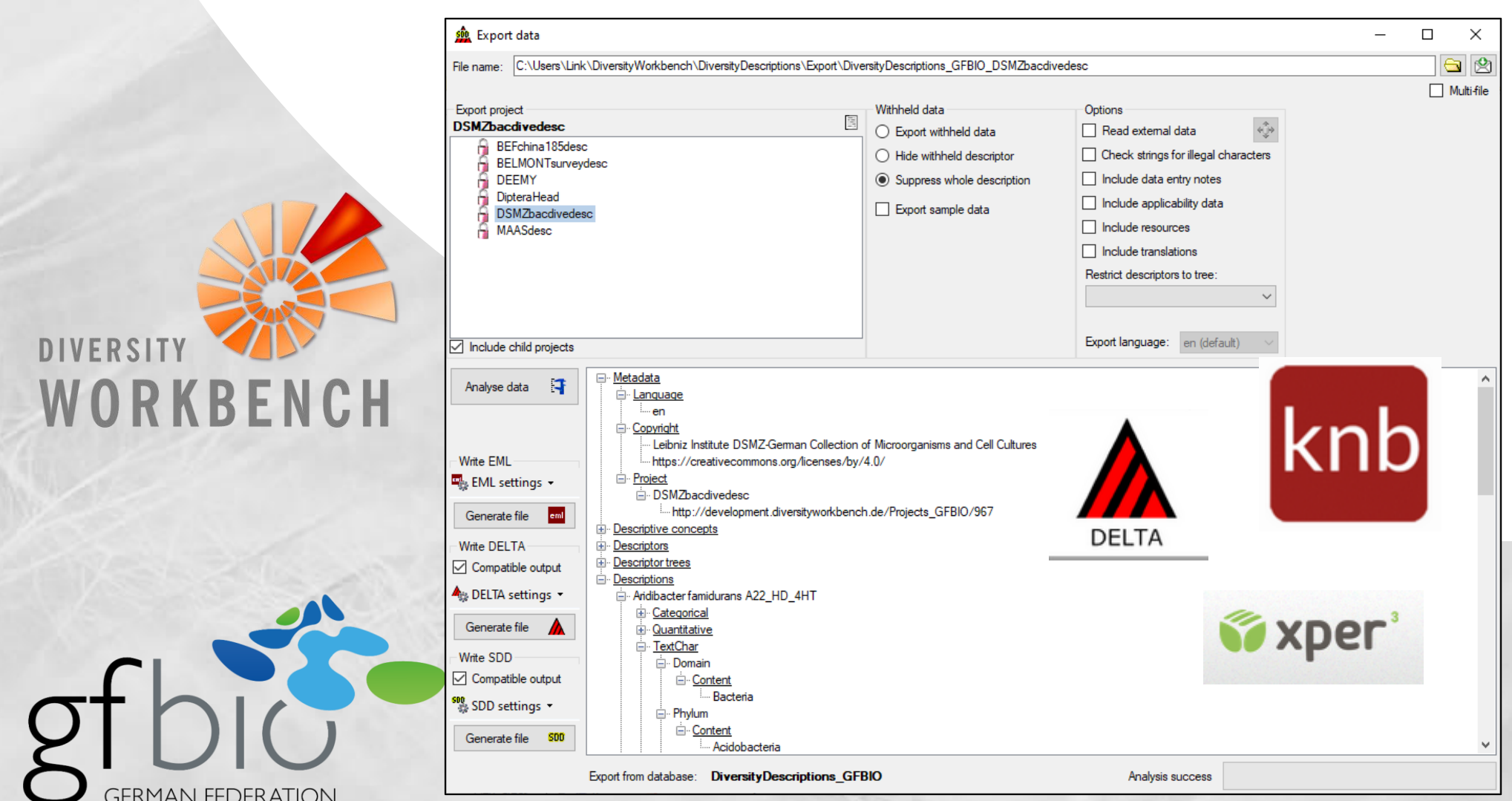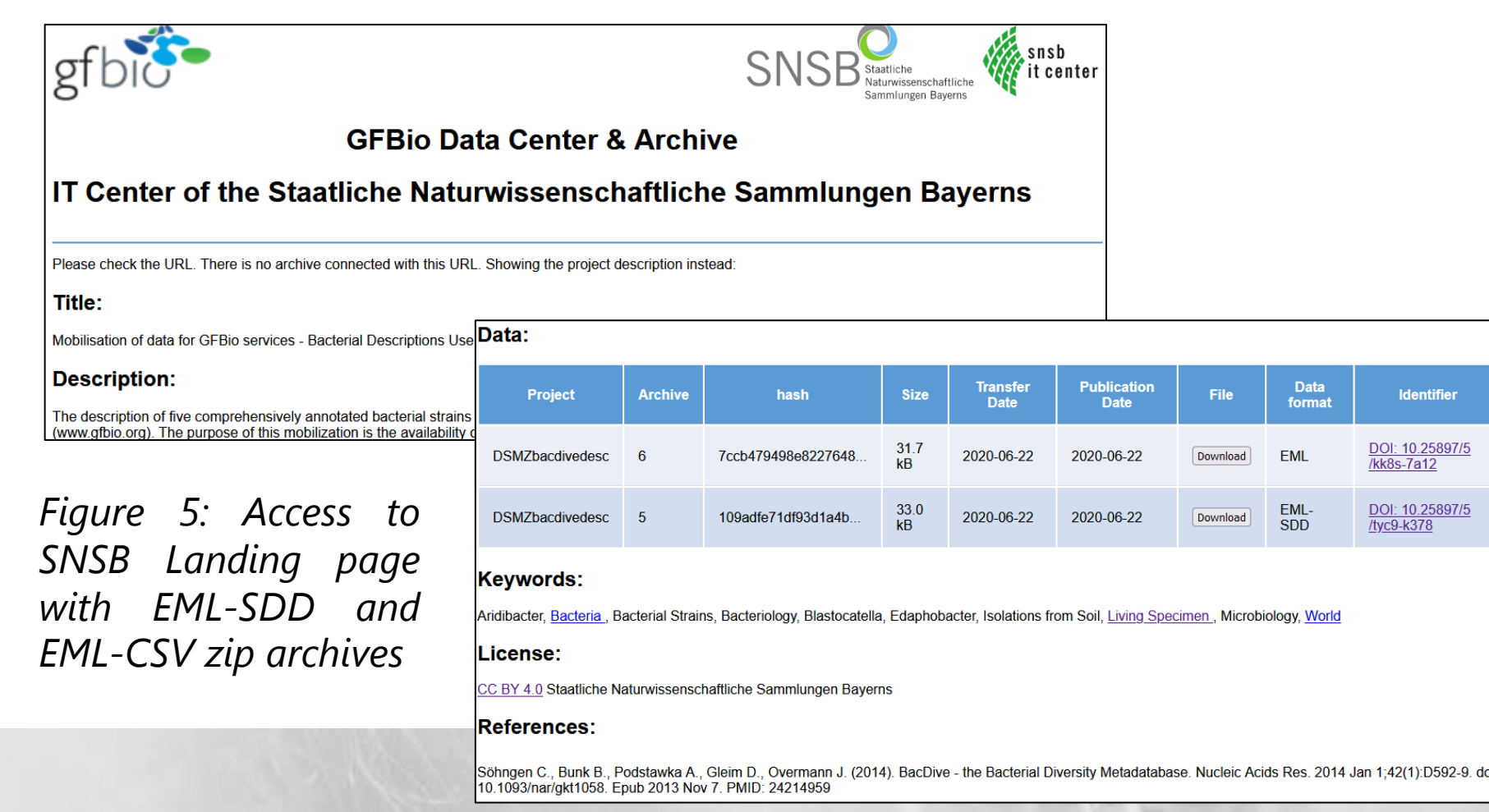


*Figure 5: Access to SNSB Landing page with EML-SDD and EML-CSV zip archives*