



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost



## Metodika NERP

# System sběru, zpracování a publikace dat pro potřeby environmentálního reportingu



CZ.03.4.74/0.0/0.0/15\_025/0016059

Optimalizace systému řízení příjmu, validace, zpracování a reporting datových sad v resortu životního prostředí (NERP)

Česká informační agentura životního prostředí

Vytvořil:	Jiří Přech, tým NERP	Datum:	16.1.2023
Schválil:	Ing. Miroslav Kukuc	Datum:	6.2.2023

Revize	Datum uveřejnění	Aktualizoval	Poznámka / Komentář
1.0	6.2.2023	Miroslav Kukuc	Publikovaná verze

## Obsah:

1. Účel a rozsah platnosti metodiky .....	3
2. Související dokumenty .....	3
3. Definice pojmů a zkratk .....	3
4. Role (Odpovědnost/pravomoci) .....	4
4.1. Poskytovatel dat .....	4
4.2. Kurátor dat .....	4
4.3. Správce dat .....	4
5. Postup při sběru, zpracování a publikace dat .....	5
5.4. Sběr dat .....	5
5.5. Zpracování dat .....	6
5.6. Publikace dat .....	6

## 1. Účel a rozsah platnosti metodiky

Účelem metodiky je upravit jednotný postup při sběru, zpracování a publikaci dat o životním prostředí pro potřeby informační podpory environmentálních politik.

## 2. Související dokumenty

### Obecně závazné předpisy

Zákon	Zákon č. 106/1999 Sb., o svobodném přístupu k informacím
Zákon	Zákon č. 123/1998 Sb., o právu na informace o životním prostředí
Usnesení	Usnesení vlády ČR č. 446/1994 a 934/2014

### Vnitřní předpisy CENIA

Strategie	Strategie národního reportingu dat o životním prostředí
Metodika	Metodika národního reportingu dat o životním prostředí
Směrnice	SME 10-2022 Pravidla publikování v CENIA

## 3. Definice pojmů a zkratk

Termín	Popis
Poskytovatel dat	Pracovník spolupracující organizace.
Kurátor dat	Pracovník Oddělení informační podpory environmentálních politik.
Správce dat	Pracovník Oddělení zpracování a správy dat.
Datová sada	Datová tabulka, která představuje základní jednotku v datovém toku.
Instance datové sady	Katalogový záznam datové sady, který vzniká k příslušnému roku a obsahuje aktuální data a metadata.
Definice datové sady	Základní sada informací, které se jednotným způsobem skladují pro účely využití v návazných procesech (kód datové sady, název, názvy sloupců, specifikace datových typů, přiřazení k číselníkům, příznaky, metadata, šablona a další technické podklady).
Šablona datové sady	Standardizovaná struktura datové sady, která se v běžném případě nemění v čase. Tato struktura je v případě každé datové sady předem vytvořena na straně CENIA, a to správcem datové sady ve spolupráci s jejím kurátorem.
Metadata	Soubor údajů popisující datovou sadu, zejména její věcný obsah, časové, územní a další souvislosti.
Indikátor	Indikátory vznikají zpracováním datových sad pro účely hodnocení stavu a vývoje životního prostředí. Představují základní jednotku reportingu. Určují obsah a rozsah publikovaných výstupů, ale i způsob ukládání dat nebo vizualizace.
Dataport	Vstupní portál CENIA nabízející rozhraní pro ukládání a editaci datových sad, indikátorů, příslušných souborů, metadat a vazeb do datového skladu a řízení jejich publikace.

Katalog dat	Poskytovaná data jsou katalogizována v CKAN. Katalog slouží především pro archivaci a prohledávání datových sad a příslušných metadat.
-------------	--

## 4. Role (Odpovědnost/pravomoci)

### 4.1. Poskytovatel dat

- Odevzdává datovou sadu a příslušná metadata
- Potvrzuje podmínky užití dat
- Pracuje s instancemi datových sad
- Může udělat opravu již odevzdané datové sady

### 4.2. Kurátor dat

- Udržuje seznam definic datových sad (spolu se správcem):
  - zakládá nové datové sady, přiřazuje k nim název, kód a kontaktní osoby (správce, poskytovatel)
  - edituje existující definice
  - udržuje datové sady aktuální, popř. je označuje jako neaktivní
  - udržuje aktuální metadata v definici
  - zakládá a udržuje podmínky užití a publikovatelnost DS jako otevřená data
  - zakládá instance datových sad
  - podílí se se správcem na tvorbě šablony DS
- Posílá žádosti o odevzdání datové sady poskytovateli
- Schvaluje odevzdané datové sady. Kontroluje vyplněná metadata, odevzdaný excel a vytvořené csv. Datovou sadu buď schválí nebo vrátí k opravě.
- Udržuje seznam definic indikátorů:
  - vytváří nové indikátory
  - udržuje jejich definice
  - udržuje indikátory aktivní / neaktivní
- Vytváří instanci k vybrané definici indikátoru
- Vytváří / aktualizuje vizualizaci indikátoru
- Edituje publikovatelnost instancí indikátoru
- Je oprávněn vkládat instance datové sady, podobně jako poskytovatel

### 4.3. Správce dat

- Správa definic datových sad – má přístup k definicím datových sad vytvořeným kurátorem:
  - vytváří a edituje jejich šablony
  - edituje názvy a datové typy sloupců
  - napojení na číselníky
  - zpracovává nutné vstupy pro další části systému (např. ETL skript, validační skript)
- Poskytuje podporu v případě technických potíží na vstupu
- Spravuje definice indikátorů
- Spravuje instance indikátorů
- Publikuje indikátory jako otevřená data

## 5. Postup při sběru, zpracování a publikace dat

### 5.4. Sběr dat

- Identifikace informačních potřeb a specifikace požadavků: Kurátor, na základě schválených publikačních plánů MŽP, identifikuje potřebu datových sad pro zpracování indikátorů a osloví poskytovatele dat.
- Smluvní zajištění poskytování datových sad: CENIA uzavře Smlouvu o poskytování dat s organizací (Poskytovatel dat). Ve smlouvě jsou definovány poskytované datové sady, včetně formátu a způsobu dodání a podmínek užití dat.
- Technické zajištění datových sad a příprava šablon:
  - Datová sada je tematicky homogenní, z hlediska svých dimenzí úplná datová tabulka, která představuje základní jednotku v datovém toku. Z praktického hlediska jejího zpracování odpovídá:
    - jednomu listu v xls datového souboru odevzdávaného poskytovatelem
    - jednomu csv souboru vytvořenému z xls souboru
    - jedné faktové tabulce v databázi
  - Pro každou datovou sadu je evidováno její unikátní id (kód datové sady), standardizovaný název a definovaná struktura – schéma datové sady.
  - Minimálním požadavkem pro jakoukoli datovou sadu, kterou CENIA obdrží od poskytovatelů, je předem stanovená, standardizovaná struktura (schéma datové sady), která se v běžném případě nemění v čase. Tato struktura je v případě každé datové sady předem vytvořena správcem datové sady ve spolupráci s jejím kurátorem a prokonzultována s poskytovatelem dat.
  - Ke každé datové sadě lze jednoznačně přiřadit koherentní sadu metadat dle aktuální Metodiky národního reportingu dat o životním prostředí.
  - Součástí definice datové sady je šablona, která splňuje všechny potřebné standardy a maximálně usnadňuje automatizaci zpracování. Za šablonu je po obsahové stránce odpovědný kurátor. Po formální stránce definuje šablonu správce, ten zabezpečuje soulad se strukturou vhodnou pro strojové zpracování. Součástí procesu je identifikace a mapování číselníků.
- Odevzdání datové sady je realizováno poskytovatelem na dohodnutém rozhraní a v dohodnutém termínu. Poskytovatel je informován o datových sadách k odevzdání. Primárním způsobem odevzdání datových sad je Dataport (vstupní portál):
  - Přístup k aplikaci pro poskytování datových sad je zajištěn na stránce <https://dataport.mzp.cz/>.
  - Dataport nabízí rozhraní pro ukládání a editaci datových sad, indikátorů, příslušných souborů, metadat a vazeb do datového skladu a řízení jejich publikace.
  - Přístup do aplikace je povolen pouze uživatelům s přiděleným oprávněním, které administruje CENIA.
- Validace formální správnosti datových sad probíhá automaticky při odevzdání dat přes vstupní portál, nicméně součástí validace je také manuální kontrola kurátorem.

## 5.5. Zpracování dat

- Katalogizace dat
  - Přijímaná data jsou katalogizována v CKAN. Katalog slouží jednak pro prohledávání datových sad a příslušných metadat a zároveň jako úložiště souborů pro nahrávání do datového skladu.
  - Indikátory jsou rovněž katalogizovány a příslušné soubory (embed kód vizualizace, data a metadata, příznaky pro publikaci) jsou přebírány webovým portálem.
- Uložení dat
  - Pro ukládání dat slouží datový sklad. Data se z katalogu nahrávají do stage vrstvy a následně jsou modelována v tabulárním (analytickém) modelu. Datové sady se dekomponují na faktová data a dimenze.
  - Relační databáze umožňuje analytickou práci s dimenzemi napříč datovým modelem.

## 5.6. Publikace dat

- Definice a výpočet indikátorů: Indikátory vznikají zpracováním datových sad pro účely hodnocení stavu a vývoje životního prostředí a jsou publikovány v rámci indikátorových sad periodických hodnotících publikací a strategických dokumentů. Indikátory tvoří „základní jednotku“ reportingu v tom smyslu, že určují obsah a rozsah publikovaných výstupů, ale i způsob ukládání dat nebo vizualizace.
- Vizualizace indikátorů se vytváří v BI nástroji nad analytickým modelem. Výsledné reporty (vizualizace) jsou živě připojené na databázi. Při změně dat (aktualizace, opravy) se změny přímo promítnou do publikovaných vizualizací.
- Data v publikacích: Procesy a datové workflow podporují přípravu každoročních hodnotících publikací (Zpráva o stavu životního prostředí ČR, Zprávy o životním prostředí v krajích ČR, Statistická ročenka životního prostředí ČR), vyhodnocení indikátorových sad strategických dokumentů (Hodnocení zranitelnosti ČR ve vztahu ke změně klimatu, Indikátory státní politiky životního prostředí) a přípravu dalších speciálních publikací či monografií.
- Publikace dat na webovém portálu: Webový portál slouží k prezentaci indikátorů a datových sad obsažených v katalogu CKAN směrem k veřejnosti. Svůj obsah přejímá portál z katalogu.
- Publikace otevřených dat: LKOD přebírá identifikované indikátory, datové sady a jejich metadata z datového skladu pro Národní portál otevřených dat (NKOD).