

# Databáze mrtvých webových zdrojů - Technická dokumentace

---

Specializovaná veřejná databáze **Databáze mrtvých webových zdrojů** zobrazuje ve strukturované formě údaje o zaniklých webových zdrojích archivovaných Webarchivem Národní knihovny ČR. Databáze je integrovaná do kurátorské aplikace **Seeder** a zobrazuje údaje získané prostřednictvím aplikace **Extinct Websites**. Obě aplikace byly navrženy jako open source, zdrojové kódy a dokumentace jsou volně dostupné v repozitářích na githubu. Informace o webových zdrojích, které jsou aplikací Extinct Websites identifikovány jako mrtvé, jsou prostřednictvím API přenášeny do Seederu a prostřednictvím **Databáze mrtvých webových zdrojů** jsou interpretovány veřejnosti - <https://www.webarchiv.cz/mrtve-weby>.

## Hlavní komponenty specializované databáze

- **Seeder** - kurátorská aplikace a webové stránky Webarchivu
- **Extinct Websites** - aplikace pro automatizované řešení identifikace a popisu mrtvých webových zdrojů

## Seeder

Databáze mrtvých webových zdrojů je integrovaná do kurátorské aplikace Seeder, která slouží pro správu webových zdrojů, licenčních smluv, rejstříku vydavatelů webových stránek, harmonogramu sklizní a správu webových stránek Webarchivu NK ČR. Aplikace byla vyvinuta na míru potřebám Webarchivu. Je napsána v programovacím jazyku Python s využitím frameworku Django. Pro uložení svých dat využívá databázi PostgreSQL.

Data z Extinct Websites se pravidelně nahrávají do aplikace Seeder pomocí vlastního REST API rozhraní. Data se následně zobrazují uživatelům aplikace ve formě statistik, tabulek a interaktivního grafu, který využívá knihovnu Chart.js.

## Technická dokumentace pro seeder

- Github: <https://github.com/WebarchivCZ/Seeder>
- Wiki: <https://github.com/WebarchivCZ/Seeder/wiki>
- ReadTheDocs: <https://seeder.readthedocs.io/en/latest/>

## Extinct Websites

Databáze mrtvých webových zdrojů zobrazuje údaje získané prostřednictvím aplikace Extinct Websites, která slouží jako automatizované řešení pro identifikaci a popis mrtvých webů. Aplikace data ukládá do vlastní interní databáze a zpřístupňuje je kurátorům, kteří s informacemi v ní dále nakládají, interpretují je a obsah klasifikují. Aplikace Extinct Websites identifikuje mrtvé weby za pomoci stavových kódů, dle kterých weby kategorizuje na skupiny, jimiž jsou automatizovány další procesy, jako je ověřování metadat z živých webů, databáze WhoIS, či historických metadat.

### Techická dokumentace pro Extinct Websites

- Github: <https://github.com/WebarchivCZ/extinct-websites>
- Wiki: <https://github.com/WebarchivCZ/extinct-websites/wiki/Popis-aplikace>

### Popis API

Základní popis api je na [wiki Extinct Websites](#)

Pro použití s Databází mrtvých webových zdrojů je důležité přiřadit parametru **type** na hodnotu **seeder**, tedy:

```
http://url-aplikace/api/v2/?type=seeder
```