

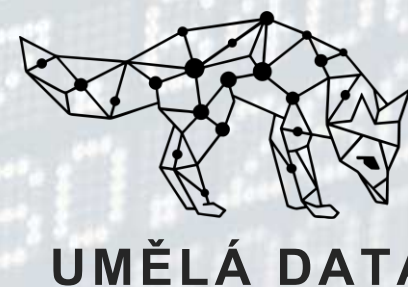
Umělá data

jako nástroj bezpečného sdílení dat

ISSS, 13.05.2025

JN ANALYTICS

ING. JIŘÍ NOVÁK, PH.D.



Sdílení dat ve veřejné správě

-  **Rostoucí tlak na otevřená data a transparentnost**
 - veřejnost, výzkumníci i politici žádají přístup k detailním datům
-  **Omezení daná ochranou osobních údajů**
 - riziko prozrazení citlivých informací
-  **Napětí mezi užitečností a ochranou**
 - čím lépe chráněná data, tím méně použitelná
-  **Tradiční anonymizace má své limity**
 - u složitějších struktur často selhává

Co jsou umělá data?

- Umělá data jsou synteticky vytvořená data, která neobsahují žádné skutečné osoby ani subjekty, ale zachovávají statistickou logiku a strukturu původních dat.
- **Co je cílem?**
 - Využít data bez ohrožení soukromí
 - Umožnit bezpečné analýzy, vývoj, publikace
- **Klíčové znaky:**
 - ❌ Žádné reálné údaje o jednotlivcích
 - ✅ Zachování statistických vlastností a vztahů v datech
 - 🤖 Vytvářeno pomocí AI/ML/statistických modelů

Jak se umělá data tvoří?

-  **Analýza původních dat**
 - identifikace proměnných, vztahů a rozdělení
-  **Výpočet modelu**
 - pomocí AI, ML nebo statistických modelů
-  **Generování nových záznamů**
 - vytvoření nové datové sady bez reálných údajů
-  **Ověření kvality a soukromí**
 - kontrola, zda data zachovávají statistické vlastnosti a neobsahují rizikové záznamy

PŮVODNÍ DATA

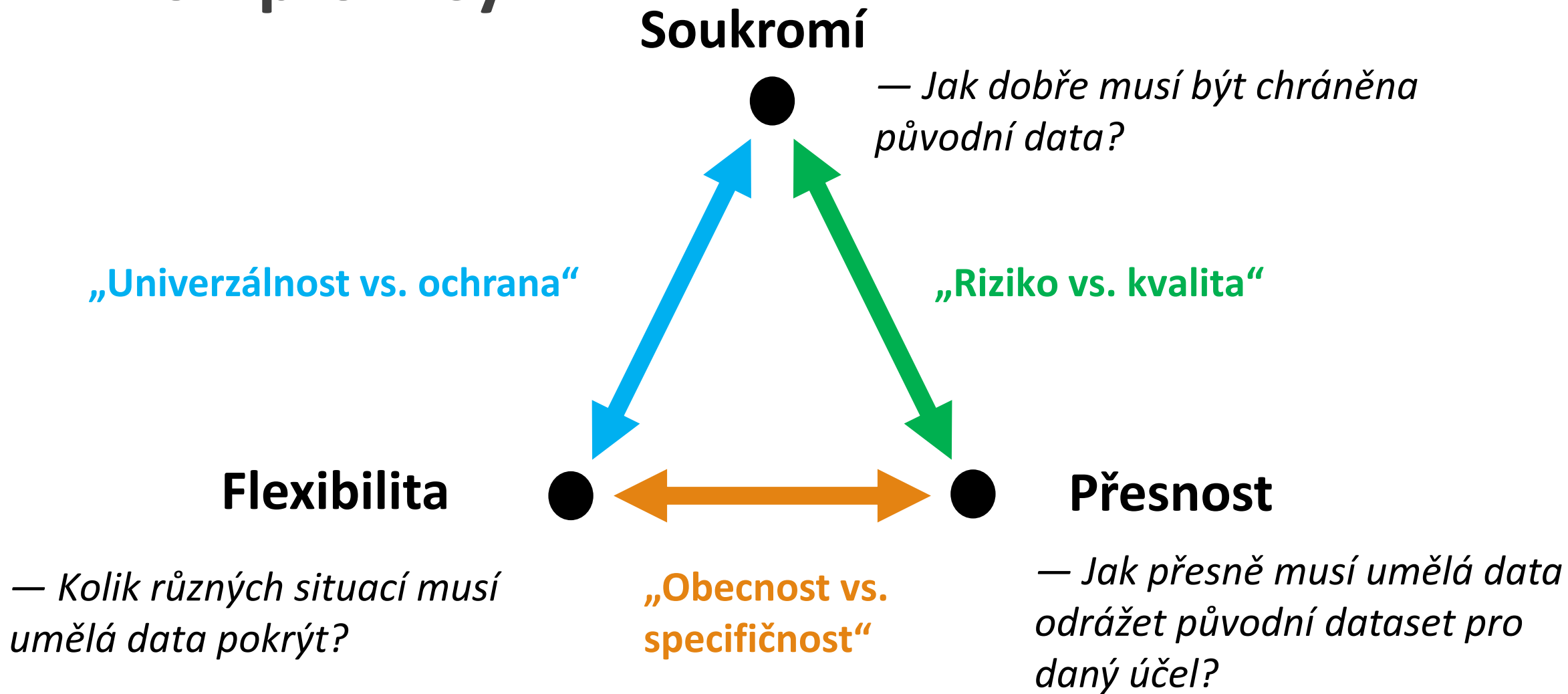


MODEL



UMĚLÁ DATA

Kompromisy



Příklady využití ve veřejné správě

-  Český statistický úřad (ČSÚ)
 - Využití umělých dat při ochraně mikrodat ze Sčítání lidu
 - Zajištění detailních dat pro výzkum bez ohrožení respondentů
-  Ústav zdravotnických informací a statistiky (ÚZIS)
 - Syntetická data ze zdravotnických registrů (NZIS)
 - Pro přípravu výpočetních skriptů spouštěných pod kontrolou ÚZIS
-  Potenciál pro další instituce
 - Otevřená data pro samosprávy, výzkum, digitalizaci agend
 - Simulace, testování, školení na realistických, ale bezpečných datech

Proč (a kdy) používat umělá data?

-  **Bezpečné sdílení dat**
 - bez rizika narušení soukromí nebo důvěrnosti
-  **Testování a školení**
 - realistické scénáře bez přístupu k originálním údajům
-  **Otevřenost a transparentnost**
 - zpřístupnění komplexních dat veřejnosti, výzkumníkům či firmám

**Umělá data nejsou náhradou za realitu,
ale bezpečnou cestou, jak ji lépe zpřístupnit
a přitom chránit to, co chráněno být musí — naše soukromí.**



Ing. Jiří Novák, Ph.D.



www.jnanalytics.cz



www.umeladata.cz



jiri.novak@jnanalytics.cz



www.linktr.ee/jiri_novak



UMĚLÁ DATA



Přízemí - Stánek 52