

M U N I
I C S

**Podrobnější představení implementace
ukládání dat
v jednotlivých CEITEC CF**

Tomáš Svoboda

Jak to začalo?

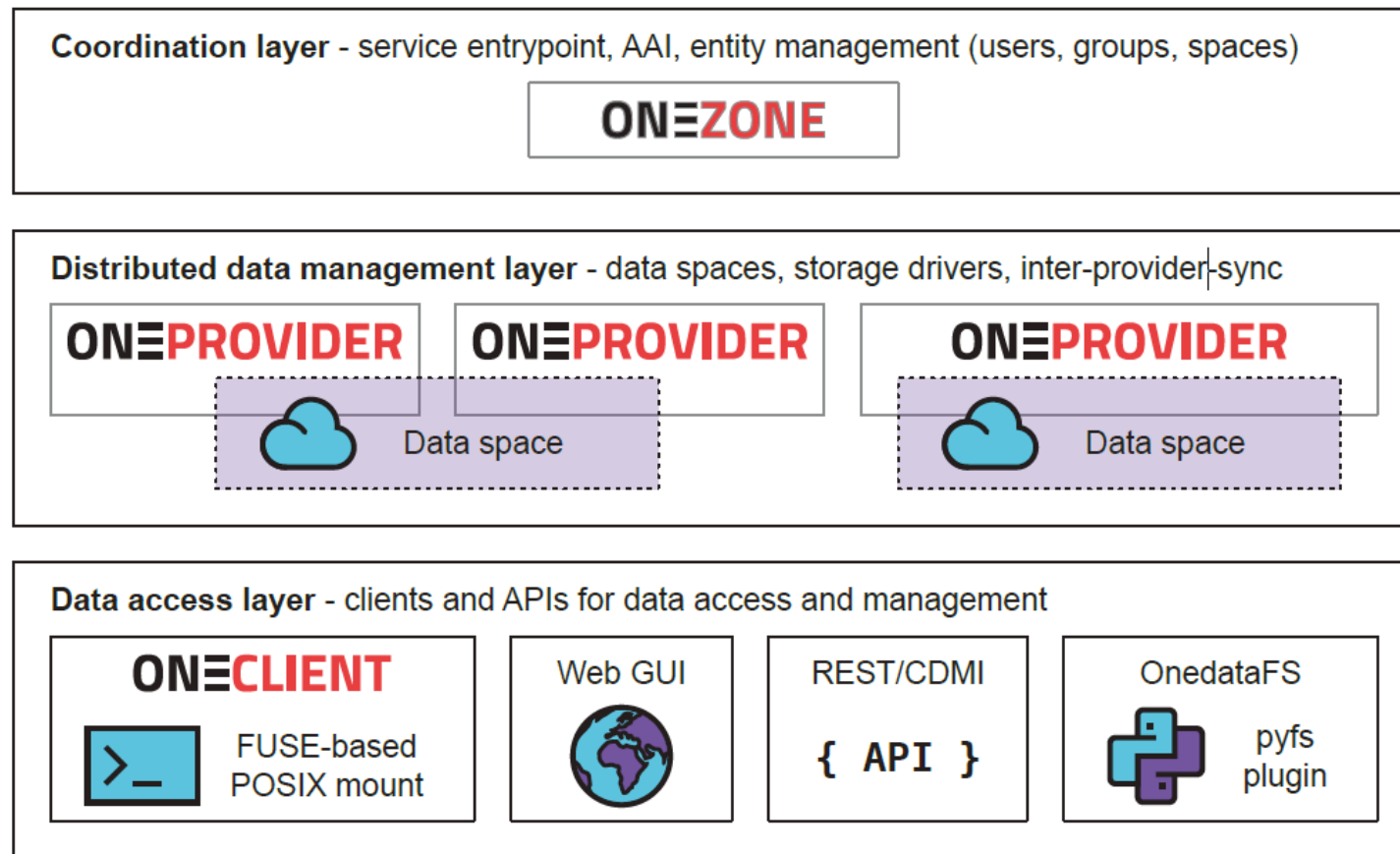


ONE DATA



Onedata

- System pro globální správu a přístup k datům
 - ukládání, sdílení, archivace a publikace
- Přístup k datům jako známá cloudová řešení (Dropbox, ...)
- Uzpůsobeno pro vědecké prostředí
 - Podpora FAIR principů
 - Podpora HPC prostředí
- Úložné technologie
 - POSIX, S3, Ceph, ...



Zkušenosti

- 🔍 Negativní proroci
- 🔍 Instalace
- 🔍 Dokumentace
- 🔍 Navázání kontaktu s autory

The screenshot displays the EGI DATAHUB Onezone interface. The left sidebar contains navigation options: Overview, Data, Shares, Transfers, Providers, Members, and Harvesters. Below these are several dataset entries with their sizes and share counts. The main panel shows the 'DATA' view for the selected dataset, '220321_ribosome_glutaraldehyde', provided by MUNI-ICS. It lists files with columns for 'Files', 'Size', and 'Modification'.

Files	Size	Modification
Gain	—	23 Mar 2022 10:56
Movies	—	23 Mar 2022 11:15
SPA.yml	3.2 KiB	29 Mar 2022 12:57

Aplikace fs2od

- Automatizovaný workflow životního cyklu datové sady
- Konfigurovatelné možnosti
 - Načítání metadat
 - E-mailové notifikace
 - Replikace (přiblížení dat k výpočtům)
 - Archivace (pro dlouhodobé uložení)
 - Mazání (z primárního úložiště)



Frontend



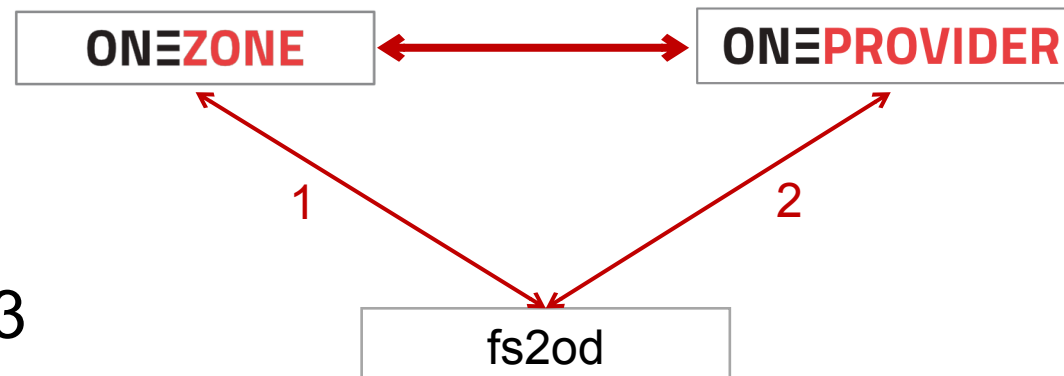
Backend



**Realita nebyla
tak růžová**

Zádrhel #1

- 🔗 Vzrůstající pravděpodobnost výskytu chyby při alokaci úložiště pro datovou sadu (300+)
- 🔗 Debuggování
- 🔗 Nedůvěra, ze strany autorů
- 🔗 Problém při interní komunikaci v rámci Onedata
- 🔗 Nelineární složitost ověřování konzistence dat v Onezone



- 🔗 Fix leden 2023

Zádrhel #2

- Náhodné zaseknutí registrace existujících souborů (+1000)
- Space ID (8cd781e98253ccb7f632b1b952314937ca2097)
 - „Náhodně generované“
 - V interním plánovači se pracuje jen 8 B prefixem, což je ale ve skutečnost 4 B
 - Při kolizi zastavena akvizice dat pro všechny datové sady
 - Nástroj tmate

- Fix srpen 2023

Zádrhel #3

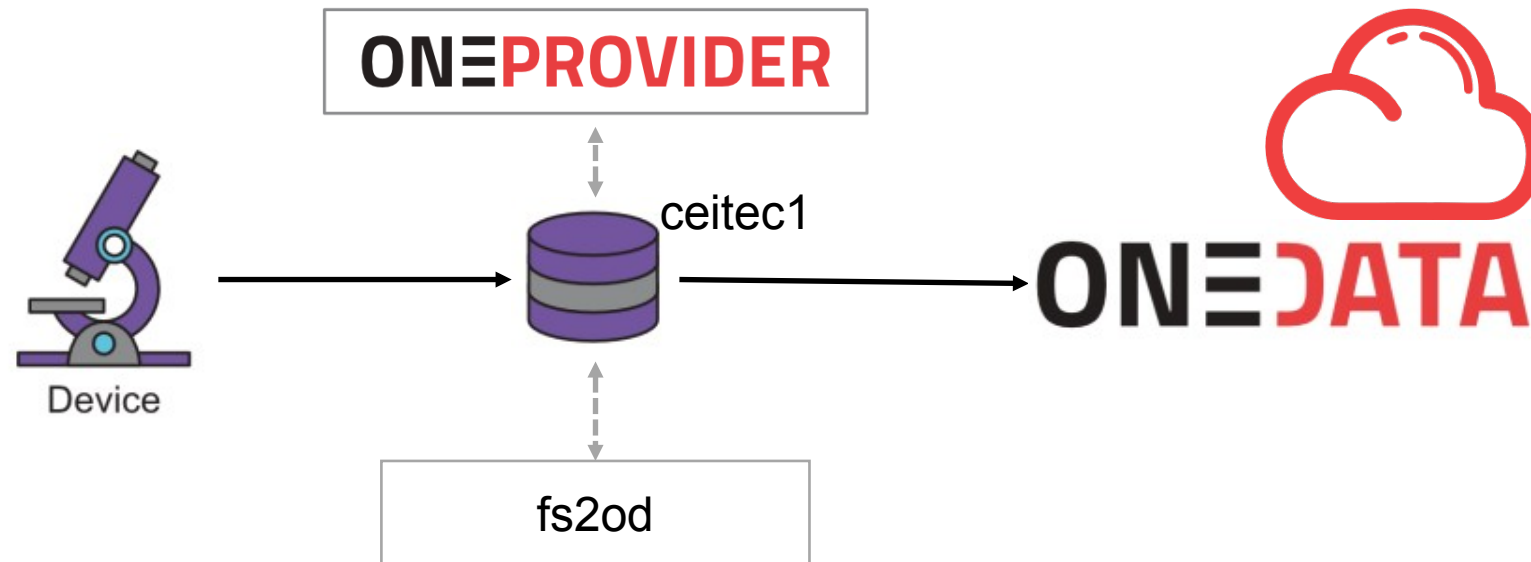
- 🔗 Webové GUI nepřipravené pro tak velké množství datových sad (800+)
 - 🔗 Pracujeme přes API, ale GUI se používá při ručních zásazích a kontrole
 - 🔗 Při velkém počtu datových sad selže načtení celé stránky
-
- 🔗 Fix květen 2023 (částečně)

CEMCOF



CEMCOF

- 🔗 Vlastní velké diskové pole, Oneprovider přímo na čelním uzlu

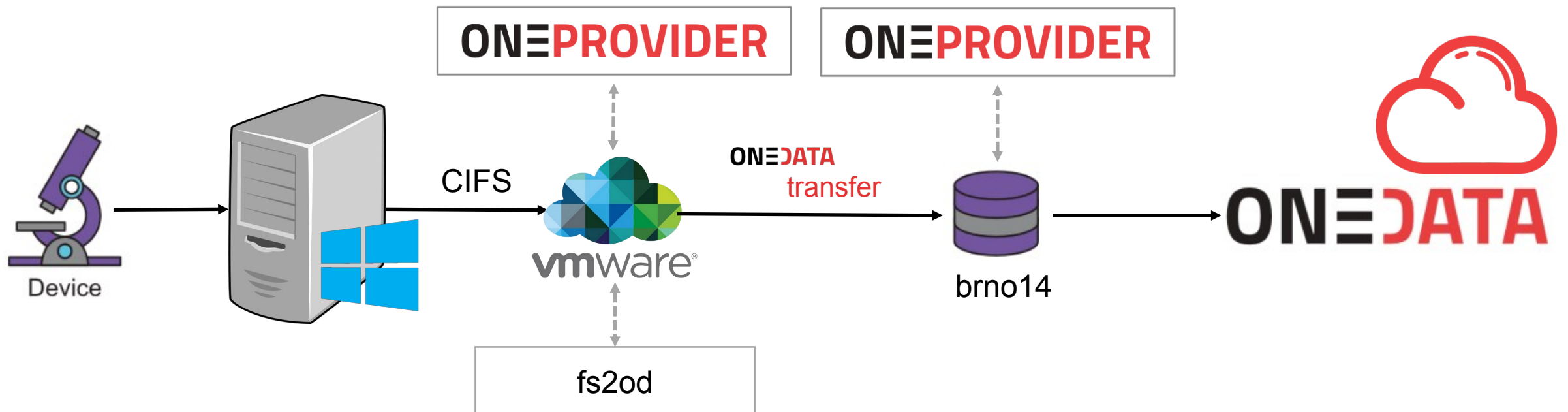


CF CELLIM

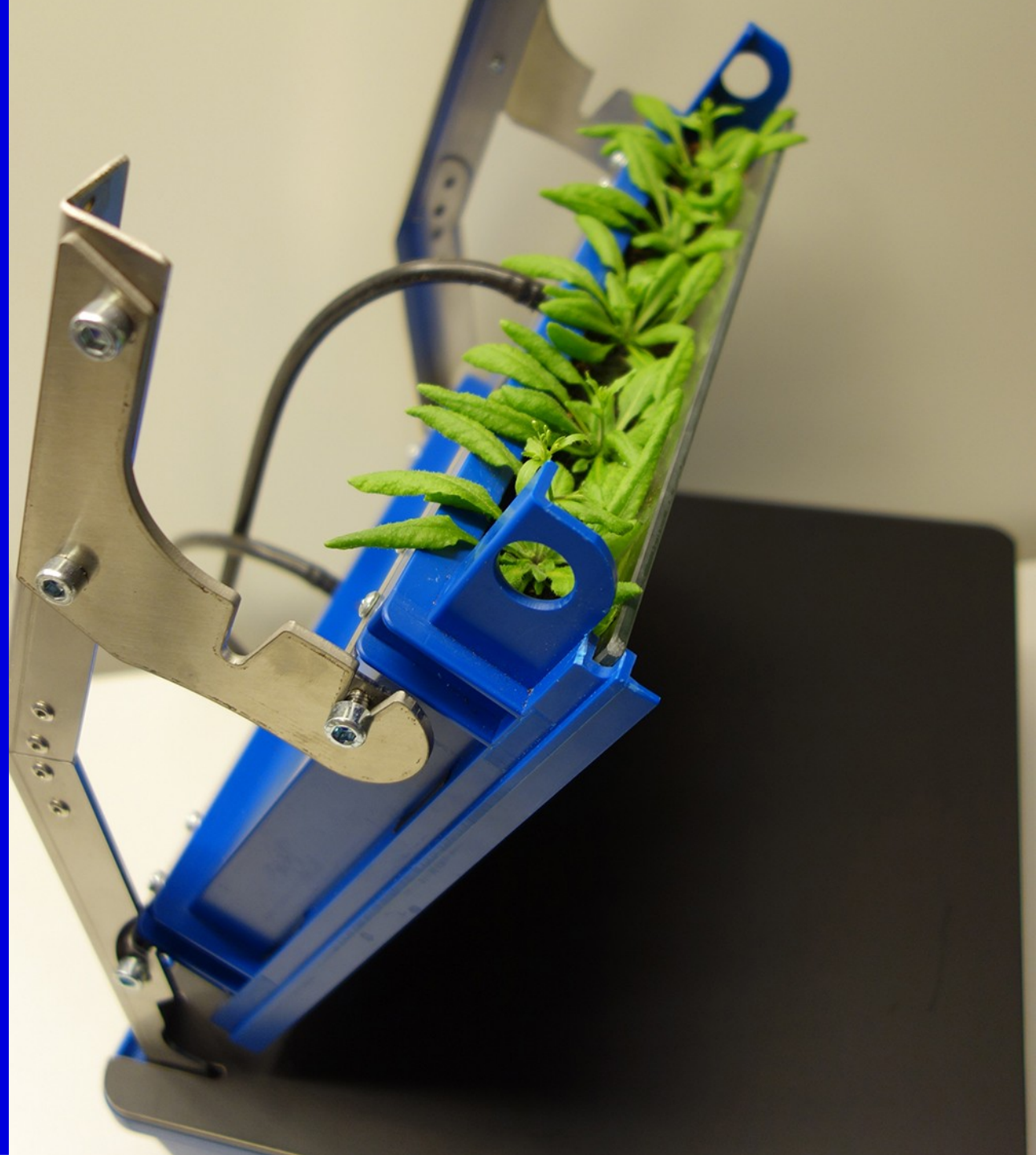


CF CELLIM

- Windows Server
- Zpracování a přechodné uložení dat

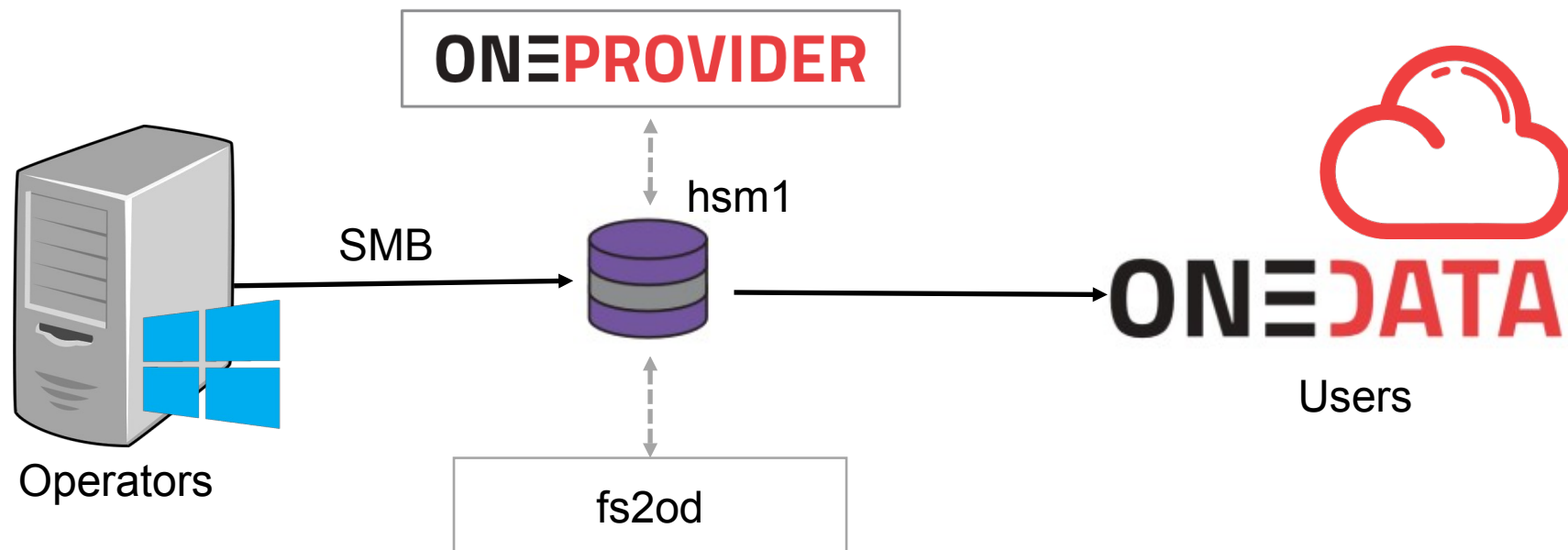


CF PLANTS

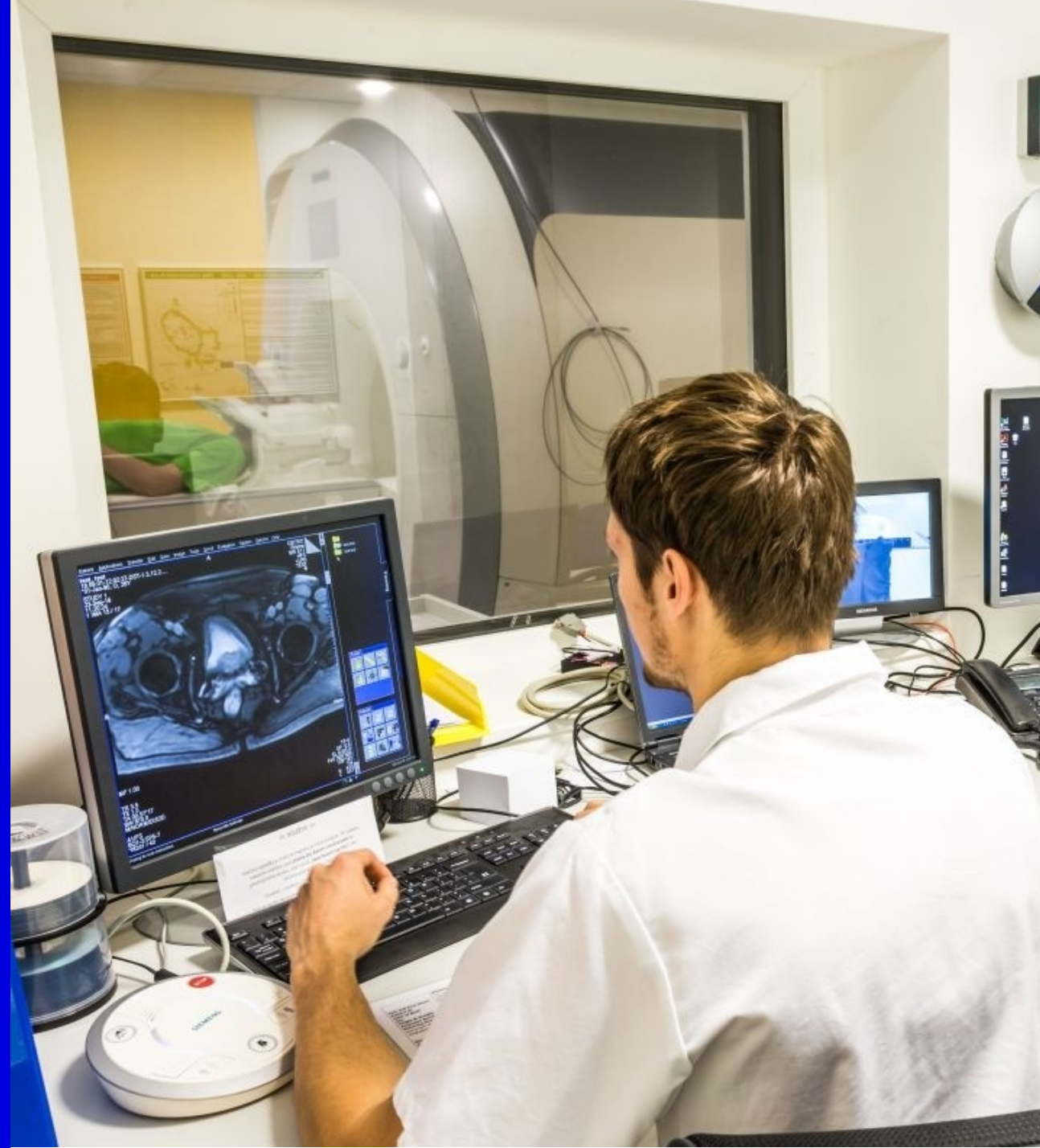


CF PLANTS

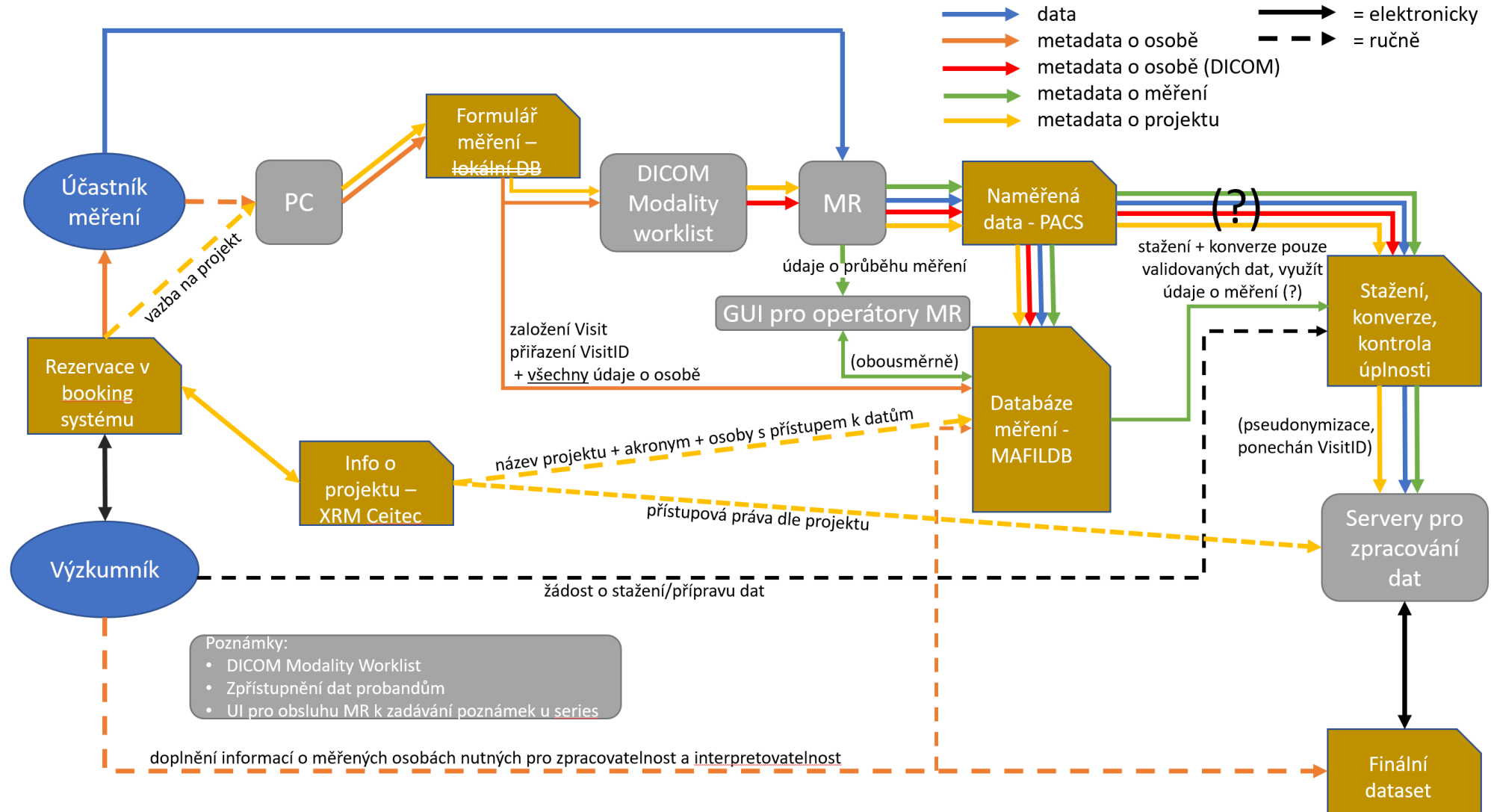
- 🔗 Předání výsledků uživatelům
- 🔗 Úvahy o repozitáři



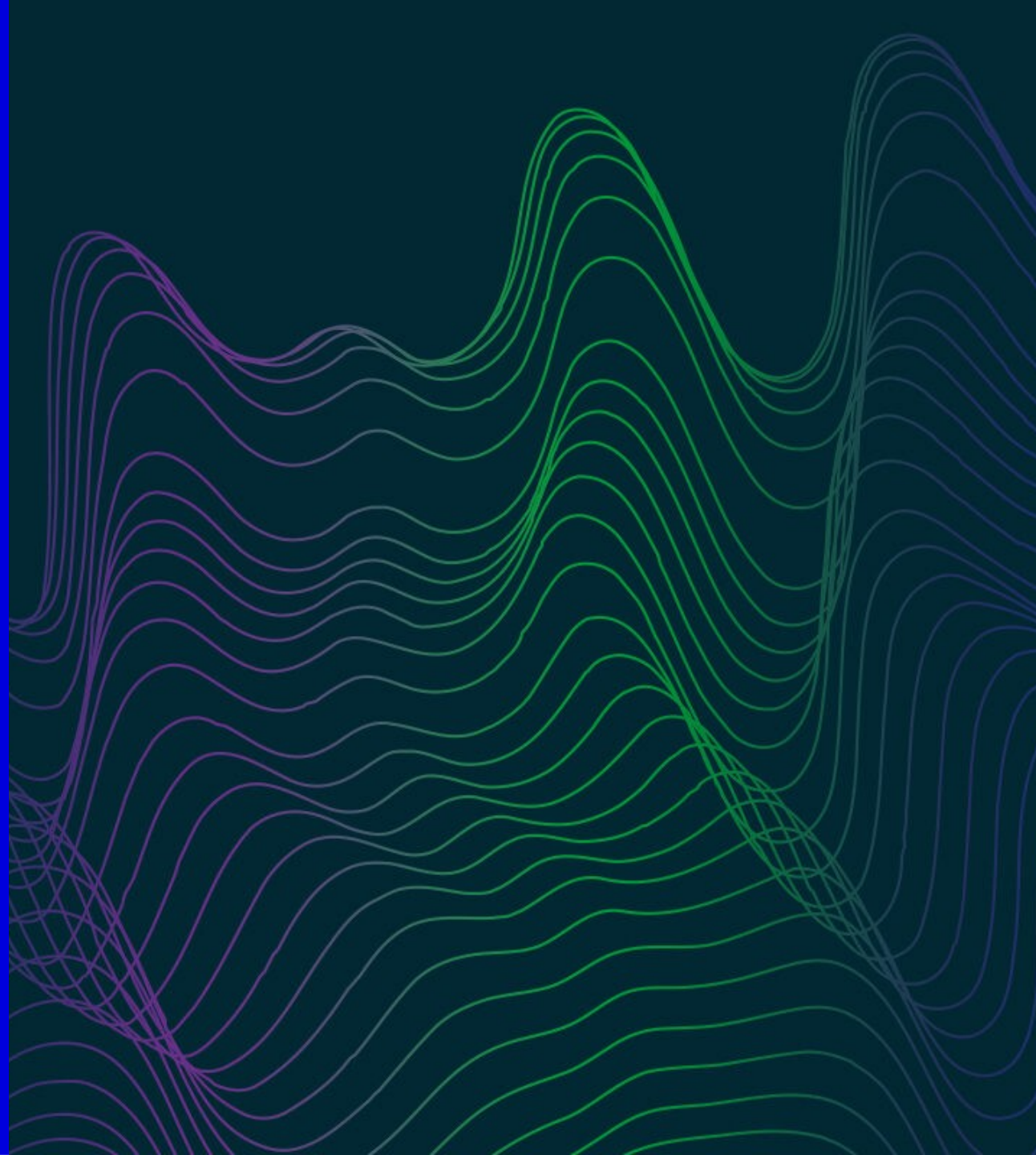
CF MAFIL



CF MAFIL



Projekt RELIEVE



Projekt RELIEVE

- 🔍 Výzkum neurologických nemocí
- 🔍 Návrh datového workflow (EEG data)

