

Směrnice S2023-01

Příloha č. 2

Podrobná specifikace pro archivaci primárních dat a laboratorních deníků pro jednotlivé obory provozované v ÚOCHB

Integrita výzkumu a správa dat

Výzkum v ÚOCHB by se měl provádět podle nejvyšších standardů integrity výzkumu. Zásadní otázkou je náležitá správa dat.

Klíčovou a nedílnou součástí prostředí pro správu dat je používání laboratorních deníků, kam se zaznamenává každý experiment, a systém inventarizace nových sloučenin a jejich charakterizačních údajů. Laboratorní deník může být buď (nejlépe) elektronický, nebo vázaný papírový sešit, do něhož se záznamy zapisují nesmazatelným perem. Každý experiment musí být datován.

U všech publikovaných prací (včetně recenzovaných prací i rukopisů deponovaných v archivních serverech a databázích nebo bakalářských, diplomových a disertačních prací) je třeba uchovávat primární údaje, které jsou základem publikovaných prací, jako doklad jejich autenticity a pro umožnění reprodukovatelnosti v souladu se zásadami správy dat FAIR. Všechna primární nebo nezpracovaná data by měla být archivována a bezpečně uložena po dobu minimálně 10 let od prvního zveřejnění na nejméně dvou nezávislých paměťových médiích (pevné disky, CD/DVD atd.).

Výše uvedené údaje by měly být na odůvodněnou žádost zpřístupněny. Pokud jsou data předmětem embarga nebo je přístup k nim omezen, měla by být poskytnuta alespoň základní metadata.

CHEM klastr

Klastr chemicky zaměřených skupin (CHEM) ÚOCHB pokrývá organickou syntézu, bioorganickou a medicínskou chemii i materiálovou chemii. Všechny nové sloučeniny je třeba plně charakterizovat analytickými metodami a spektroskopií (NMR, MS, IR atd.) a následně zaznamenat do elektronického inventárního systému. Všechna nezpracovaná NMR data se musí archivovat v původním formátu FID spolu s akvizičními parametry; archivovány by měly být i nestandardní parametry zpracování. Surová MS data je nutno uložit ve formátu a softwaru poskytnutými dodavatelem spektrometru. Ostatní surová spektra by se rovněž měla archivovat ve formátu poskytovatelů přístrojů. V případě rentgenové krystalografie se krystalografické informační soubory (CIF), tabulky strukturních faktorů a validační zprávy CheckCIF předávají Krystalografickému datovému centru v Cambridgi (Cambridge Crystallographic Data Centre – CCDC). Všechny stabilní nové sloučeniny (i když nejsou určeny pro medicínskou chemii) by měli vědečtí pracovníci předávat do knihovny sloučenin ÚOCHB k testování biologické aktivity.

BIO klastr

Klastr biologicky zaměřených skupin (BIO) ÚOCHB pokrývá molekulární buněčnou biologii, strukturní biologii a chemickou biologii. Proteomická, lipidomická a transkriptomická data je třeba ukládat do příslušných databází (PRIDE atd.), jak obvykle vyžadují časopisy, nebo archivovat v ÚOCHB.

Další primární data, která se zpravidla nezařazují do sekcí rozšířených nebo doplňkových dat, mohou zahrnovat např. primární neořezané imunoblots, snímky imunofluorescence (světelné mikroskopie), surová data o rychlostech enzymových reakcí použitá pro měření IC50, další primární enzymatická a vazebná data atd. Pro všechny gelové elektroforézy je nutno archivovat původní skeny gelů v surovém formátu (např. RAW). Vystává závažný a zřetelný problém surových dat z experimentů kryoelektronové mikroskopie (cryo-EM), která u jediného experimentu s vysokým rozlišením mohou snadno dosáhnout desítek TB. Režim zálohování dat tohoto typu a přístupu k nim se bude muset s největší pravděpodobností lišit od ostatních dat (např. pomocí magnetických pásek).

PHYS klast

Uchovávání dat z výpočetních prací zajišťuje reprodukovatelnost a integritu vědeckých výsledků a dalšího výzkumu na nich založeného. Za účelem zajištění reprodukovatelnosti získaných výsledků by měli vědci uchovávat alespoň všechny vstupní soubory/data, kódy a instrukce k jejich provedení. Mezi tyto soubory a podrobné informace patří výchozí struktury a podmínky simulace, databáze a podobné informace, které mohou získané výsledky přímo a podstatně ovlivnit. Uchovávané informace by měly rovněž zahrnovat úplnou identifikaci použitých kódů/softwaru (např. verze, možnosti kompilace, parametry); interní kódy nebo skripty by se měly uchovávat pod licencí open-source. Je nutné podporovat používání plně anotovaných protokolů, které obsahují všechny informace zajišťující reprodukovatelnost postupů (např. zápisníky Jupyter). Ačkoli to není vždy proveditelné (především kvůli problémům s velikostí), je žádoucí podporovat uchovávání získaných nezápracovaných dat, zejména těch, která byla použita v publikované práci.