



Plán správy dat

Ústav geoniky AV ČR, v. v. i.

1. Definice „dat“

Za „data“ relevantní k tomuto plánu považujeme pouze ty datové sady, které:

- 1) musí být zpřístupněny podle pravidel a principů otevřeného přístupu k vědeckým datům (dále „otevřená data“);
- 2) data, která se povinně zpřístupňovanými daty podle aktuálních pravidel mohou stát v budoucnu, a
- 3) data vytvořená v zájmu mimoústavních subjektů v rámci smluvního výzkumu nebo činnosti infrastruktur.

Všechna ostatní data shromážděná nebo zpracovaná během výzkumné činnosti Ústavu geoniky AV ČR, v. v. i., (dále „ÚGN“) se považují za interní data, která jsou buďto nedůležité povahy (tj. nemusí být zpřístupněna pro podporu závěrů výzkumu), nebo se jedná o data, pro která existuje důvod zabráňující zveřejnění, tedy platná výjimka.

Vzhledem k multioborovému zaměření výzkumné činnosti ÚGN vzniká převážná většina výzkumných dat ve velmi rozmanitých a specifických přístrojových platformách, přičemž mnohé z nich jsou vlastní konstrukce. Jejich relevance pro externí zájemce je tedy značně omezená a porovnatelnost a reprodukovatelnost na pracovištích mimo jejich původ může být velmi limitovaná.

Formáty dat, s nimiž je pracováno na ÚGN, mohou být velmi různorodé, což reflektuje široký výzkumný záběr pracoviště. Data lze kategorizovat rámcově následovně:

1. Data tradičních formátů zpracovatelná běžným, široce rozšířeným software. Do této kategorie spadají číselné a textové soubory, obrazová data v běžných formátech aj.
2. Data získaná proprietárním software, jenž je součástí speciálních, ale komerčních přístrojů provozovaných na ÚGN.
3. Data zaznamenaná přístroji konstruovanými na ÚGN a ukládaná ve formátu specifickém pro konkrétní aparaturu.
4. Data konstrukčního charakteru.
5. Anonymizovaná data získaná dotazníkovým šetřením.

Využití dat v interním prostředí a jejich správa pro potřeby výzkumu řešitelských kolektivů jednotlivých projektů je ve správě výzkumných oddělení, týmů a řešitelů.

2. FAIR data

2.1. Zajištění dohledatelnosti dat (Findability), včetně ustanovení pro metadata

Metadata jsou nezbytnou součástí ukládaných datových souborů a nesou informace o době vzniku datového souboru, okolnostech jeho záznamu, typu experimentu a projektu, v jehož rámci řešení byla pořízena. Metadata související s jejich interpretací mohou v řadě případů zahrnovat prvky duševního vlastnictví a jejich zveřejnění či předání externím subjektům podléhá pravidlům pro zpřístupnění dat s aplikačním potenciálem (viz níže).

Vzhledem k širokému oborovému rozkročení ÚGN jsou další detaily struktury metadat specificky vázány na zvyklosti jednotlivých příslušných oborů. V tomto plánu správy dat nejsou a ani

nemohou být jmenovitě upravována již jen pro variabilitu jednotlivých projektů a jejich specifických požadavků.

Data ÚGN nemají (až na výjimky) charakter dat sbíraných na souborech objektů či vzorků, které by měly relevanci pro externí strojové zpracování (data harvesting). Jejich úprava pro strojové čtení by nebyla smysluplná a mohla by vést k nežádoucí desinterpretaci.

2.2. Zpřístupnění dat (Accessibility)

Formy zpřístupnění dat ÚGN lze kategorizovat podle charakteru subjektu, kterému mají být k dispozici:

1. Zpřístupnění pro řešitelský kolektiv ÚGN

Zpřístupnění dat pro interní potřeby řešitelských kolektivů projektů řešených na ÚGN je plné a neomezované. Přístup do datových úložišť je chráněn standardními obvyklými postupy zabezpečení dat. Metadata odpovídají interním potřebám řešitelského kolektivu.

2. Zpřístupnění pro pracovníky ÚGN mimo řešitelský kolektiv konkrétního projektu

Zpřístupnění dat pro osoby mimo konkrétní řešitelský kolektiv v rámci ÚGN je plné jako v bodě 1, pokud není řešitelský kolektiv smluvně vázán mlčenlivostí (NDA).

3. Zpřístupnění pro členy konsorcia projektů, s nimiž ÚGN váže smlouva o spolupráci

V případě dat vzniklých v rámci projektů, které zahrnují externí spoluřešitele, jsou formy ukládání dat, charakteru metadat a formy sdílení dat určeny smlouvami s partnery, případně zadávacími dokumentacemi jednotlivých konsorciálních projektů. Tyto projektové a smluvní požadavky jsou v konkrétních projektech nadřazeny tomuto plánu správy dat.

4. Zpřístupnění pro externí subjekty

Data, která jsou identifikována jako data, pro něž je zpřístupnění externím subjektům umožněno, jsou zpřístupněna v režimu “upon reasonable demand” – na základě smysluplného požadavku externího subjektu. Pro žádost o poskytnutí dat lze využít formuláře na webových stránkách ÚGN. Požadavek je poté zhodnocen osobou Data Stewarda (viz níže), který jej předá relevantnímu pracovníkovi ústavu. Rozhodnutí, jedná-li se o smysluplný (“reasonable”) požadavek, náleží řešiteli projektu, v jehož rámci data vznikla. V případě potřeby řešitel konzultuje s vedoucím týmu, oddělení nebo Data Stewardem.

Zpřístupnění dat externím subjektům není anonymní, požadavek na předání dat je vázán na vyplnění dotazníku zahrnujícím jméno, afiliaci a e-mailovou adresu žadajícího. Vlastní žádost musí být podána způsobem, který umožňuje ověření totožnosti žadatele, tedy datovou schránkou nebo doporučeným dopisem. Na žádosti podané emailem nebude brán zřetel.

Formy zpřístupnění dat ÚGN lze kategorizovat také podle charakteru projektu, v jehož rámci vznikla, a jeho výstupů:

1. Zpřístupnění dat souvisejících s publikacemi

Data související s publikacemi podléhají pravidlům vydavatelů a jsou zpřístupňována náležitou formou, vydavateli požadovanou.

2. Zpřístupnění dat vzniklých v rámci projektů různých poskytovatelů

Zadávací dokumentace a pravidla poskytovatelů účelové podpory v podobě projektů obvykle stanovují vlastní pravidla práce s daty. Pokud tato pravidla nejsou v souladu s plánem správy dat ÚGN, avšak ředitel ÚGN schválil podání projektové žádosti, pro konkrétní jednotlivé projekty jsou pravidla poskytovatelů účelové podpory nadřazena tomuto plánu.

Vzniklá data jsou zpřístupňována v okamžiku, kdy jsou dokončeny logické celky výzkumu, a je zřejmé, co je výsledkem daného celku, jak bude s výsledkem naloženo, a výsledek je v daném duchu řádně zpracován a připraven. Typicky se jedná o zpřístupnění dat po/s odesláním manuskriptu publikace, ukončení projektu apod., tak aby zveřejněná data odpovídala provedenému výzkumu a bylo zřejmé, že jejich zpřístupněním nebudou ohroženy oprávněné zájmy ÚGN či dalších subjektů na ochranu duševního vlastnictví nebo že zpřístupněním nebudou porušeny závazky vyplývající např. ze smluv s partnery z aplikační sféry.

2.3. Repozitáře (úložiště)

Důvěryhodnost datových úložišť a repozitářů je předmětem velké pozornosti a péče. Zásadním pro správu dat ÚGN jsou interní datová úložiště, která jsou součástí počítačové síťové infrastruktury ústavu. Péče o náležité zálohování, bezpečnost a spolehlivost provozu je povinností Úseku IT. Jednotlivá vědecká oddělení mají dedikován datový úložný prostor a využívají jej v režimu odpovídajícím vlastním potřebám. Další rozvoj a posilování interních datových úložišť jak co do zvyšování objemu uložitelných dat, tak co se týče posilování kybernetické bezpečnosti uchovávaných dat, jsou jednou z priorit ÚGN.

Mimo interní datová úložiště využívá ÚGN řady dalších repozitářů a úložišť. Lze je rozdělit následovně:

1. Datové repozitáře oborové, v široké míře využívané mezinárodní komunitou s důvěryhodností danou šíří svého používání ve vědě. Kupříkladu arXiv, nebo ZENODO.
2. Důvěryhodné národní repozitáře nabízené národními infrastrukturami pro vědu a výzkum, jmenovitě EOSC-CZ, KNAV-ASEP, CESNET.
3. Repozitáře vydavatelů vědeckých publikací, jejichž použití pro data související s publikacemi je pro autory povinností.
4. Specifické oborové repozitáře dedikované pro užší vědeckou mezinárodní komunitu, jejichž využití je obvykle nutné v rámci řešení oborových konsorciálních projektů.

Využití jiných datových úložišť a repozitářů je možné po náležitém prověření jejich důvěryhodnosti a se souhlasem vedoucího příslušného vědeckého oddělení a Data Stewarda.

2.4. Data

S ohledem na zaměření ústavu má řada výsledků aplikační potenciál a/nebo je vytvářena přímo ve spolupráci s partnery z aplikační sféry. Existuje zde tedy možnost budoucí komercializace, případně převodu do praxe v širším smyslu. Přístup k datům je pak omezen nebo embargován tak, aby byla zajištěna ochrana duševního vlastnictví. Zejména nebudou poskytována výzkumná data, která může konkurent použít k narušení ochrany duševního vlastnictví nebo k zabránění zavedení ochrany (např. zabránění patentování nebo udělování licencí). Znamená to tedy, že všechna výzkumná data týkající se potenciálně komerčně využitelných a aplikovatelných výstupů jsou považována za data, která nespádají do kategorie výzkumných dat, na která se vztahují principy otevřeného přístupu.

2.5. Nezpřístupnění dat – data se nezpřístupňují v těchto případech:

1. Data projektů aplikovaného výzkumu

Toto omezení se netýká dat souvisejících s publikacemi vytvořenými v rámci těchto projektů.

2. Data projektů smluvního a kolaborativního výzkumu financovaného z neveřejných prostředků
3. Data s aplikačním potenciálem a data zahrnující prvky duševního vlastnictví

Rozhodnutí, zda datový soubor má aplikační potenciál či obsahuje prvek duševního vlastnictví, náleží řešiteli projektu, v jehož rámci data vznikla, doporučuje se konzultace s vedoucím týmu či oddělení.

4. Data vázaná smlouvami o mlčenlivosti (NDA)
5. Data konstrukčního charakteru – strojírenské výkresy, elektrotechnická schémata, software (s výjimkou software šířeného formou Open Source)
6. Data, která byla vytvořena proprietárním software vázaným na omezení licenčními ujednáními na tento software a související metadata
7. Data s etickými omezeními
8. Data, která narušují principy ochrany osobních údajů (GDPR)

2.6. Interoperabilita dat (Interoperability)

Interoperabilita dat je pro jejich sdílení, výměnu a opakované používání zásadní. Primárním imperativem je zajištění interoperability v rámci řešitelských kolektivů pro interní použití. Proto jsou data opatřena adekvátními metadaty a srozumitelným popisem. Formáty a standardy odpovídají zvyklostem v oboru a potřebám řešitelských kolektivů. Řešitelům projektů je doporučeno používat standardní metadata, která navrhuje The Research Data Alliance, pokud není stanoveno v zadávací dokumentaci projektů či ve smlouvách o spolupráci na projektech jinak. Pro všechny typy výzkumných dat generovaných v projektu budou používány standardní slovníky.

Interoperabilita dat relevantních k vědeckým publikacím bude z technického hlediska zajištěna uložením strojově čitelné elektronické verze zveřejněné publikace nebo konečné verze recenzovaného rukopisu přijatého ke zveřejnění do příslušných repositářů. Z hlediska právního zajištění otevřeného přístupu k výzkumným datům uloženým v těchto repositářích za podmínek poslední dostupné verze veřejné licence CC BY nebo jejího ekvivalentu.

2.7. Zajištění opakovaného použití dat (Reusability)

Interní využitelnost dat je zajištěna uložením v interních datových úložištích, uložením ve formátech používaných standardně řešitelskými kolektivy, provozovanými laboratorními aparaturami, souvisejícím softwarem a opatřením adekvátními metadaty zajišťujícími srozumitelnost.

Data otevřená k přímému externímu využití jsou data vázaná na konkrétní výstupy publikačního charakteru v příslušných repositářích. Tyto zahrnují informace o dílčích výsledcích výzkumu nebo nástrojích a instrumentech potřebných k opětovnému využití výzkumných dat či k jejich validaci. Otevřený přístup k datům bude zajištěn v těchto repositářích po splnění podmínek poslední dostupné verze veřejné licence CC BY nebo jejího ekvivalentu.

Data určená ke sdílení v rámci konsorciálních projektů jsou dostupná členům konsorcia pro využití partnery projektů. Kvalita výzkumných dat bude zajištěna etickými kodexy vědeckého pracovníka platnými v institucích konsorcia.

3. Alokace zdrojů pro správu dat

Zdroje alokované péči o data na ÚGN směřují především do investic do infrastruktury počítačové sítě a datových úložišť. Na pravidelný upgrade síťových, počítačových a datových prostředků je alokován adekvátní roční rozpočet. Nákupy investičního charakteru jsou podle potřeby zařazovány do plánu přístrojových investic ÚGN.

3.1. Lidské zdroje dedikované managementu dat

1. Správa dat na úrovni projektů je úkolem řešitelů projektů.
2. Interoperabilita dat na úrovni výzkumných týmů je svěřena vedoucím týmů a na úrovni oddělení vedoucím oddělení.
3. Péče o počítačovou síť, datová úložiště ústavu, bezpečnost dat a zálohování dat je úkolem Úseku IT.
4. Role Data Stewarda, jehož úkoly jsou definovány v kapitole "Zpřístupnění dat", je svěřena pracovníkům Úseku IT.

4. Bezpečnost dat

Bezpečnost dat ukládaných na interní datová úložiště je zajišťována Úsekem IT ÚGN. Jedná se o zabezpečení před nepřátelskými útoky, před selháním techniky a před dalšími riziky ztráty dat.

Ukládání dat na externí datová úložiště a repozitáře je zabezpečeno volbou takových úložišť a repozitářů, které jsou důvěryhodné a náležitě certifikované.

5. Etika

Data vytvářená na pracovišti ÚGN jsou data o neživé přírodě, a proto nepodléhají žádným etickým omezením. Výzkum směřující k medicínským a dalším aplikacím, kde je nutné zohlednit také etické hledisko při nakládání s daty, je vždy prováděn ve spolupráci a organizován tak, aby případné experimenty na zvířecích modelech nebo výzkum na lidech byl prováděn na spolupracujícím pracovišti, které má náležitá oprávnění a užívá potřebných postupů, nikoli na ÚGN.

Tento Plán správy dat nabývá účinnosti od 1. 1. 2024.