

Blíží se »revoluce« v ukládání a sdílení vědeckých dat. Budte u toho!

8.10.2024 - Adam Pluhař | Technická univerzita v Liberci

„Pokud naši vědci pracují v projektech hrazených z veřejných zdrojů, je jednou z podmínek zpřístupňovat nejen články, ale i nasbíraná vědecká data. Ta mají být dlouhodobě dohledatelná a přístupná, tedy uložena ve spolehlivém repozitáři. Pokud je to možné, měla by být vědci nasbíraná data opatřena co nejotevřenější možnou licencí, aby je mohli pro svůj výzkum dál využívat i ostatní výzkumníci. V konečném důsledku to zkrátí čas a sníží náklady na sběr nebo tvorbu stejných dat a na možnost jejich verifikace. Obecně to pomůže urychlit vědecký vývoj,“ říká Ludvík Steiner z Univerzitní knihovny TUL, jenž poskytuje metodickou podporu Koordinátorům otevřené vědy na Technické univerzitě v Liberci.

Čeští vývojáři nyní pracují na vytvoření Národní datové infrastruktury (NDI) a repozitářů, kam se budou ukládat data nasbíraná napříč všemi výzkumnými institucemi v ČR. K budované datové struktuře systému se nyní mohou zájemci z TUL vyjádřit během diskuse se členy týmu EOSC. Setkání proběhne v pátek 11. října od 9:00 v posluchárně G314 univerzitní budovy G. *„Nyní je zprovozněn repozitář pro všeobecná data a začínají se řešit speciální, oborové typy dat: pro oblasti bio/zdraví/potravin, materiálové vědy a technologie, data management pro AI a strojové učení, sociální vědy, fyzikální vědy, humanitní vědy a umění, environmentální vědy a citlivá data. V pracovních skupinách EOSC je z naší univerzity 10 členů. Na základě požadavků a připomínek pak sdružení CESNET jako realizátor infrastruktury zapracuje naše požadavky a připomínky. Na schůzce se účastníci dozví princip fungování NDI, repozitářů i to, co se bude od autorů očekávat. Je důležité, aby právě teď zaznělo co nejvíc připomínek, aby byl systém co nejvíce přizpůsoben potřebám našich vědců,“* říká Ludvík Steiner.

Obavy vzbuzuje ukládání citlivých dat

Evropská datová infrastruktura EOSC je určena jak pro akademickou sféru, tak průmysl. Ukládaná a zpřístupňovaná vědecká data musí splňovat FAIR principy. Zatím nevyřešenou otázkou, která zároveň vědce „leká“, je však způsob ukládání citlivých dat. Například z výzkumů, které směřují k podání patentů. *„Většinu kvantitativních dat by mělo být možné zreplikovat, zároveň se tím ale dostáváme na hranu otevřené vědy, protože řada výzkumů je určena ke komercializaci nebo míří k patentovému řízení. V ten moment musí být data chráněná a někteří vědci si je drží tak říkajíc pod pokličkou. Tady nastává velká třecí plocha a nejtěžší moment celého procesu. EOSC se snaží vysvětlit, že i tato data mohou v repozitářích být, jen je zapotřebí nastavit potřebnou úroveň přístupnosti a zajistit jejich naprostou anonymitu,“* upozorňuje Ludvík Steiner.

Zároveň podotýká, že některá citlivá data mohou mít v repozitáři alespoň údaje o tom, že někde existují. I to může pro někoho představovat důležitou informaci. *„Navíc, pokud vám někdo během výzkumu sebere data, které není možné zpětně získat, a už se to několikrát stalo, vytvořená datová úložiště vás zachrání. Úložiště fungují na pěti místech republiky, takže je ztráta dat prakticky vyloučena. Vybudování vlastního repozitáře, jeho následná administrace, udržování a zajištění jednotného rozhraní s projektem EOSC je pro naši univerzitu finančně neúnosné, ale díky projektu Národní datová infrastruktura můžeme využít jednotné národní politiky a repozitáře,“* dodává Ludvík Steiner.

<https://tuni.tul.cz/a/blizi-se-revoluce-v-ukladani-a-sdileni-vedeckych-dat-budte-u-toho--157216.html>