

Mokslo duomenų valdymas: pagrindiniai principai

Gintarė Tautkevičienė, Ieva Cesevičiūtė
Kauno technologijos universiteto biblioteka

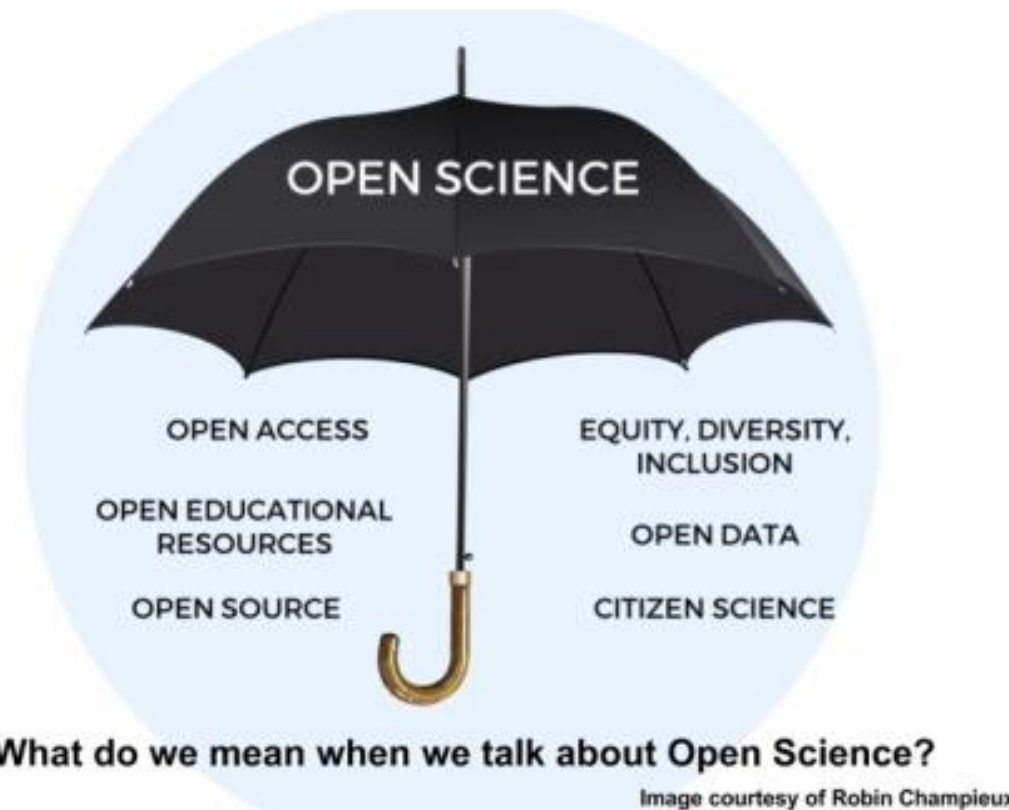
2021 m. balandžio 2 d.

Atvirasis mokslas –

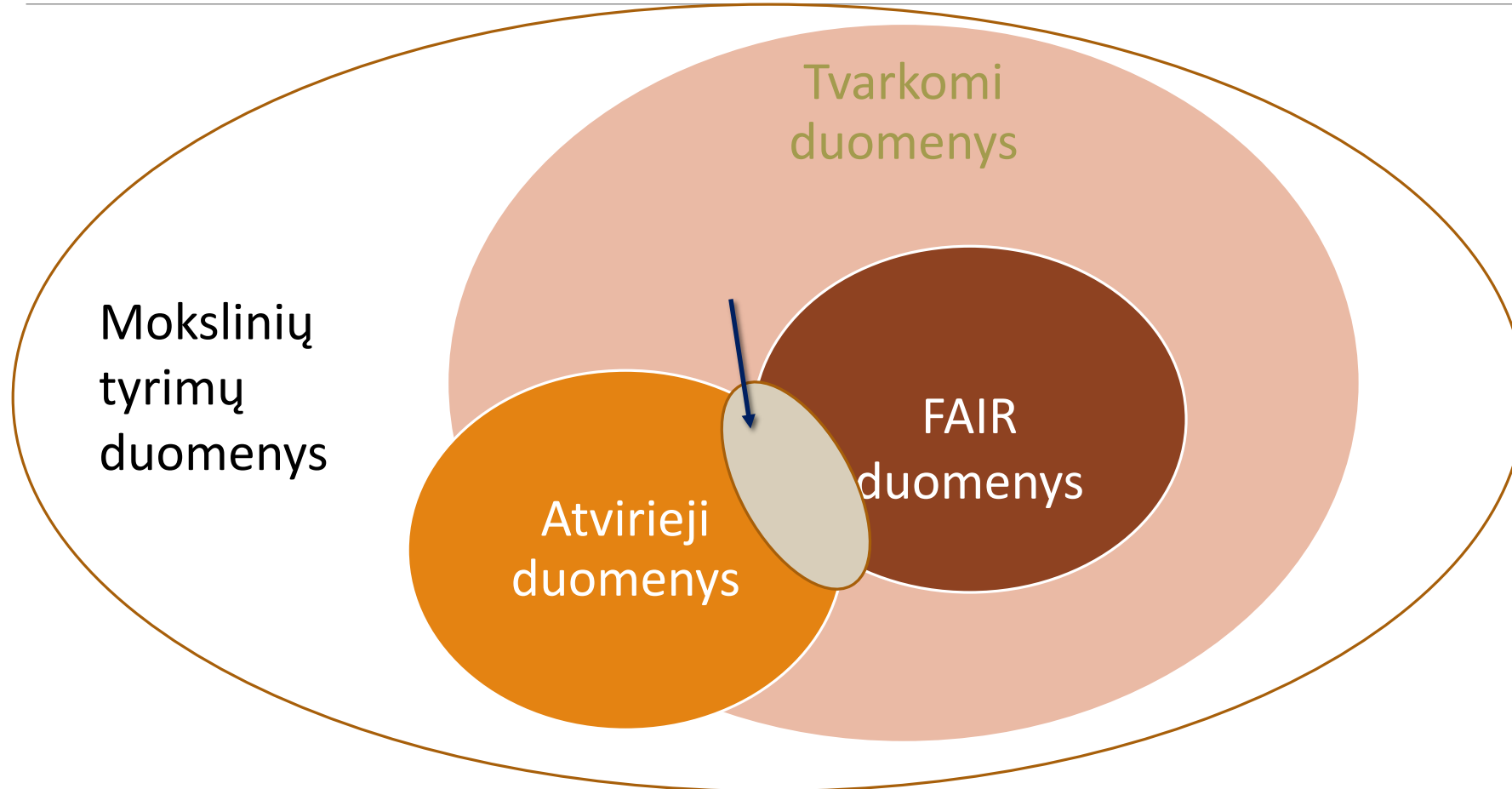
globalus judėjimas ir mokslinių tyrimų veiklos transformacija, daranti poveikį mokslinių tyrimų vykdymui ir mokslinės veiklos organizavimo metodams.

Atvirojo mokslo įgyvendinimo sąlygos:

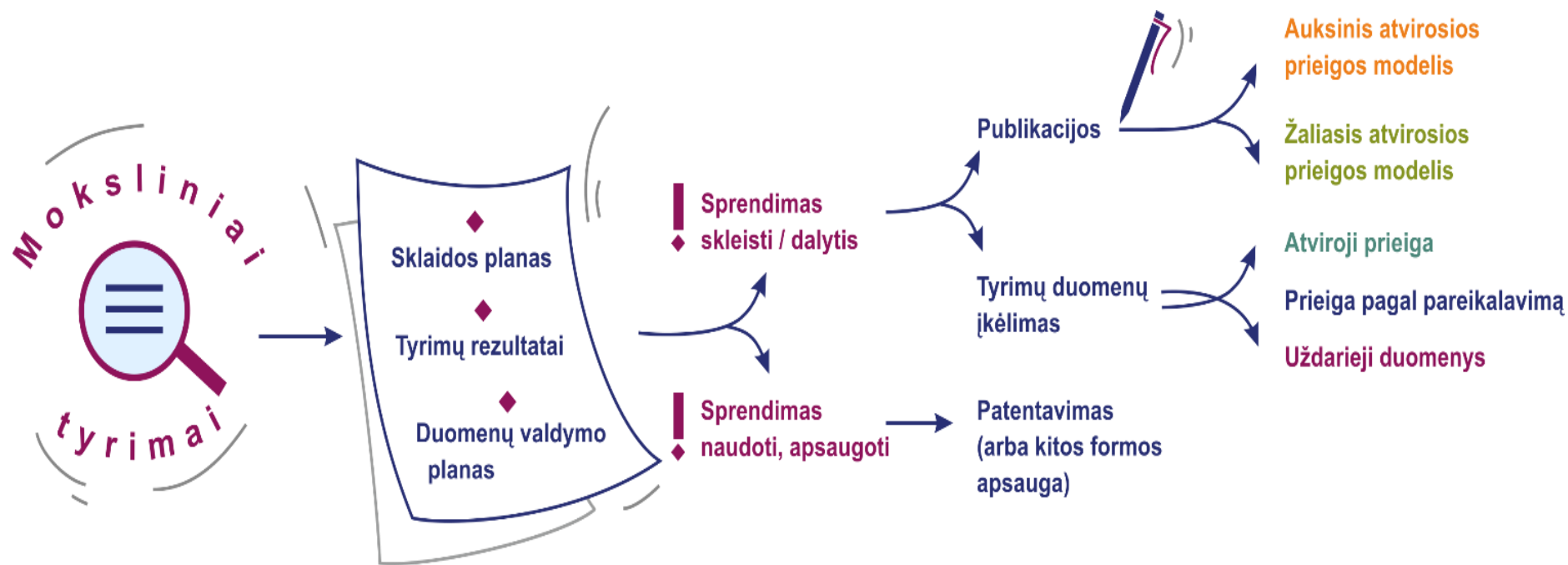
- užtikrinta atvira prieiga prie mokslinių publikacijų
- optimalus pakartotinis mokslinių tyrimų duomenų panaudojimas
- piliečių mokslas
- mokslinių tyrimų etika



Mokslinių tyrimų duomenys

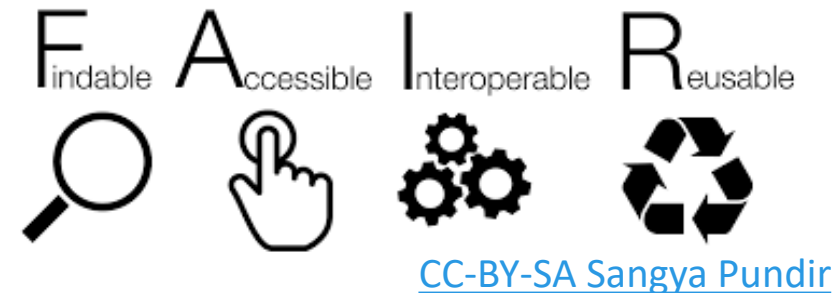


Atvirosios prieigos reikalavimai programos Horizontas 2020 projektuose



[Guidelines to rules on Open Access to Scientific Publications & Open Access to Research Data in Horizon 2020](#)
[Guidelines on Data Management in Horizon 2020](#)

FAIR duomenų principai



- Randami (angl. *Findable*)
- Prieinami (angl. *Accessible*)
- Sąveikūs (angl. *Interoperable*)
- Panaudojami iš naujo (angl. *Re-usable*)

<https://www.force11.org/group/fairgroup/fairprinciples>

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

15 FAIR duomenų principų

E- randami

F1. (Meta)duomenims suteikiamas globalus unikalus ir nuolatinis identifikatorius

F2. Duomenims aprašyti naudojami išsamūs metaduomenys

F3. Metaduomenyse aiškiai nurodomas duomenų identifikatorius

F4. (Meta)duomenys yra registruojami ar indeksuojami paieškos sistemose ar duomenų bazėse

I- sąveikūs

I1. Žinių atvaizdavimui (meta)duomenyse naudojami formalūs, prieinami, plačiai paplitę kontroliuojami žodynai

I2. (Meta)duomenyse naudojami žodynai, atitinkantys FAIR principus

I3. (Meta)duomenyse pateikiamos suderintos nuorodos į kitus susijusius (meta)duomenis

A- prieinami

A1. (Meta)duomenys gaunami pagal identifikatorius duomenų apsieitimui naudojant standartizuotus protokolus

A1.1 Protokolai yra atviri, nemokami ir universaliai diegiami
A1.2 Jeigu reikia, turi būti numatytos procedūros autorizuotam ir autentifikuotam prisijungimui

A2. Metaduomenys turi išlikti prieinami net tuomet, kai duomenys tampa neprieinami

R – pakartotinai panaudojami

R1. (Meta)duomenyse yra išsamiai aprašyti tikslūs ir tinkami duomenų atributai

R1.1. (Meta)duomenyse pateikiama aiški ir atvira duomenų prieigos licencija

R1.2. (Meta)duomenyse pateikiama detali informacija apie duomenų kilmę

R1.3. (Meta)duomenys turi būti suderinti su mokslo disciplinoje/ bendruomenėje naudojamais standartais.

2019 m. birželio 20 d. Europos Parlamento ir Tarybos direktyva (ES) 2019/1024 dėl atvirųjų duomenų ir viešojo sektoriaus informacijos pakartotinio naudojimo (nauja redakcija)

„Valstybės narės padeda užtikrinti mokslinių tyrimų duomenų prieinamumą priimdamos nacionalinę politiką ir patvirtindamos atitinkamus veiksmus, kuriais siekiama, kad viešosiomis lėšomis finansuojami mokslinių tyrimų duomenys būtų laisvai prieinami (atvirosios prieigos politika), vadovaujantis principu „atviras pagal standartą“ ir užtikrinant suderinamumą su FAIR principais. Tame kontekste į opius klausimus, susijusius su intelektinės nuosavybės teisėmis, asmens duomenų apsauga ir konfidencialumu, saugumu ir teisėtais komerciniais interesais, atsižvelgiama vadovaujantis principu „kiek įmanoma, atvira, bet, kiek būtina, uždara“. Ta atvirosios prieigos politika skiriama mokslinius tyrimus atliekančioms organizacijoms ir mokslinius tyrimus finansuojančioms organizacijoms.“ (10 str.)

Valstybės narės užtikrina, kad įsigalioję įstatymai ir kiti teisės aktai, būtini, kad šios direktyvos būtų laikomasi ne vėliau kaip nuo 2021 m. liepos 17 d. (17 str.)

Atvirieji mokslo duomenys

ATVIRA TIEK, KIEK ĮMANOMA, UŽDARA TIEK, KIEK REIKIA

Finansavimo gavėjai turi teisę atsisakyti atverti duomenis, bet privalo pagrįsti savo sprendimą



Dažniausios atsisakymo priežastys

Privatumas

Intelektinės
nuosavybės teisės

Galimas pavojus
pagrindiniams
projekto tikslams

Šis modelis buvo išbandytas vykdant programos *Horizontas 2020* bandomąjį projektą

2015 m.

Iš 431 projektų



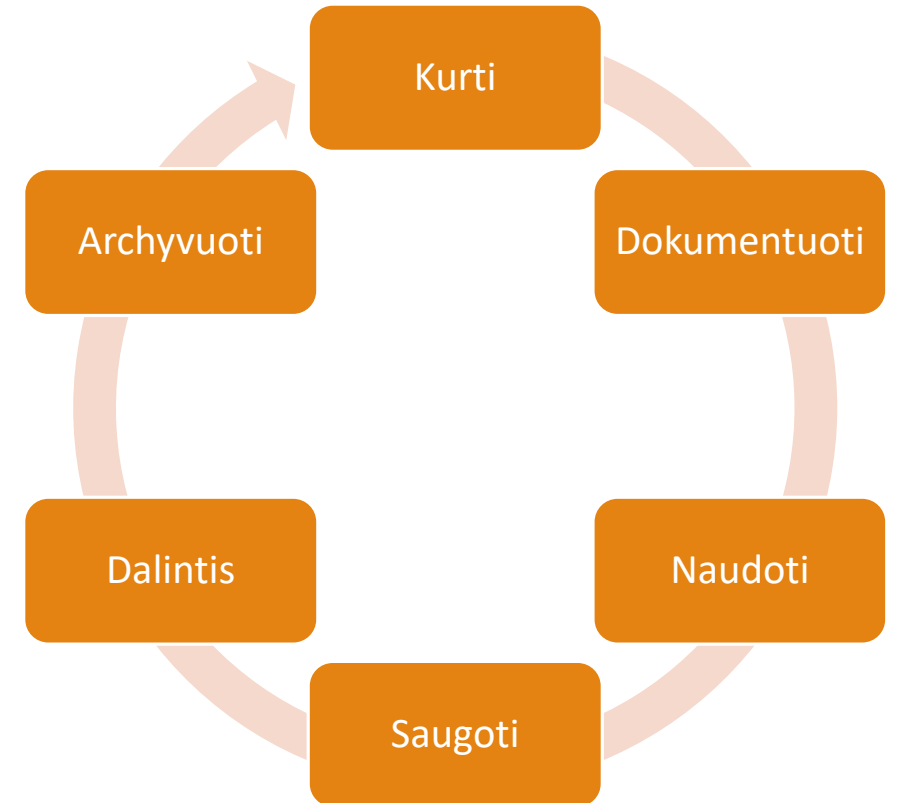
Nusprendė dalytis duomenimis

Nuo 2017 m.

Taikant tas pačias taisykles
vykdomas Atvirųjų mokslo
duomenų bandomasis projektas
apima visas programos
Horizontas 2020 mokslo sritis

Mokslo duomenų valdymas

- Apsaugo duomenis nuo praradimų, sugadinimo
- Leidžia rasti ir gauti duomenis, kai jų prireikia
- Duomenys yra saugūs
- Duomenys gali būti pakartotinai panaudojami kituose moksliniuose tyrimuose
- Leidžia dalintis duomenimis su kitais mokslininkais



Duomenų valdymo planas

Tai mokslinių tyrimų projekto duomenų kaupimo, saugojimo ir jų prieinamumo dokumentas, kuriame nurodoma:

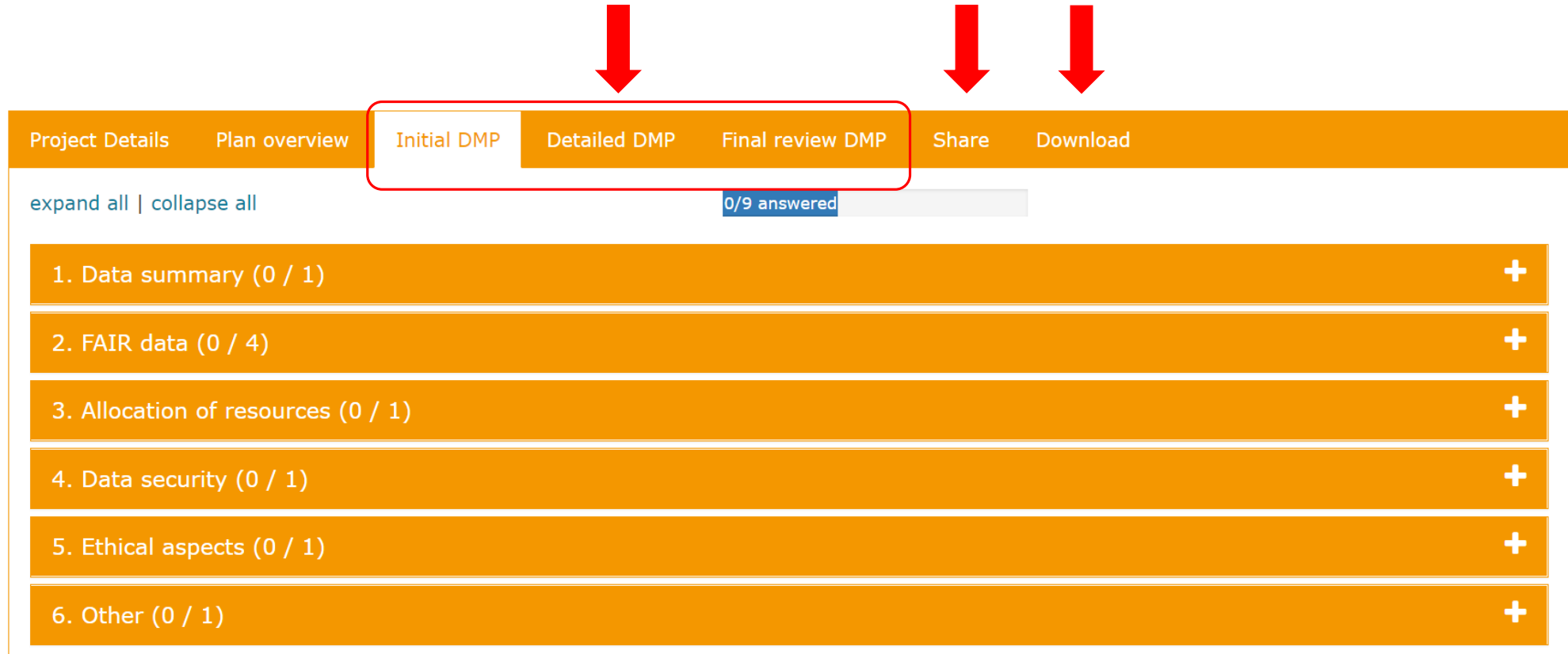
- kas bus daroma projekto metu ir jam pasibaigus, kad duomenys būtų kaupiami saugiai ir patikimai;
- kaip ir kokiomis sąlygomis duomenys bus prieinami pakartotiniam naudojimui, jei tik nėra tam prieštaraujančių teisinių, etinių ar saugumo priežasčių;
- kada ir kaip duomenys bus atveriami kitiems vartotojams;
- kokie standartiniai metaduomenys bus naudojami aprašant duomenis;
- kaip gauti duomenys bus tvarkomi ir atnaujinami;
- kurie duomenys bus skirti ilgalaikiam, o kurie – trumpalaikiam saugojimui, pastaruoju atveju nurodant, kada ir kaip duomenys bus sunaikinti.

Kodėl duomenų valdymo planas naudingas?

Duomenų valdymo planas leidžia

- prieš pradedant tyrimą apmąstyti, kaip aprašysite tyrimo procesą ir duomenis, kur juos saugosite ir kt.
- iš anksto apgalvoti ir aptarti su duomenų valdymu ir dalijimusi susijusius klausimus
- komunikuoti informaciją apie tyrimo duomenis kitiems tyrimo dalyviams, finansuojančiai institucijai ir kt. suinteresuotoms grupėms
- užtikrinti tyrimo patikimumą ir pateikti duomenis bet kada, kai jų prireikia

Duomenų valdymo planavimo įrankiai: DMPonline

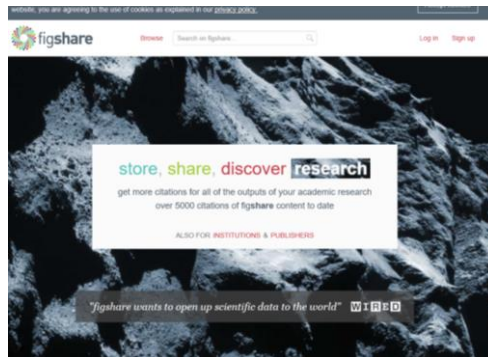


The screenshot displays the DMPonline interface. At the top, there is a navigation bar with tabs: 'Project Details', 'Plan overview', 'Initial DMP', 'Detailed DMP', 'Final review DMP', 'Share', and 'Download'. The 'Initial DMP' tab is selected and highlighted with a red box. Three red arrows point down to the 'Initial DMP', 'Detailed DMP', and 'Final review DMP' tabs. Below the navigation bar, there is a section with a list of questions and their completion status. The questions are: 1. Data summary (0 / 1), 2. FAIR data (0 / 4), 3. Allocation of resources (0 / 1), 4. Data security (0 / 1), 5. Ethical aspects (0 / 1), and 6. Other (0 / 1). Each question has a plus sign icon to its right. Above the list, there is a link 'expand all | collapse all' and a progress indicator '0/9 answered'.

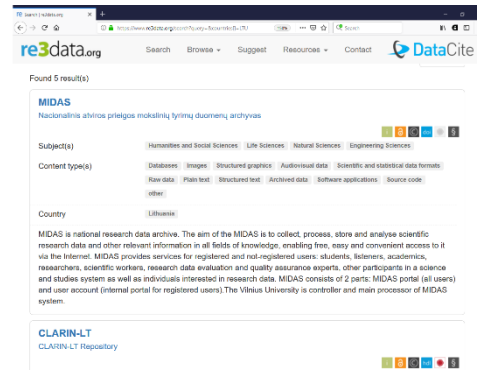
Project Details	Plan overview	Initial DMP	Detailed DMP	Final review DMP	Share	Download
expand all collapse all						
0/9 answered						
1. Data summary (0 / 1)						+
2. FAIR data (0 / 4)						+
3. Allocation of resources (0 / 1)						+
4. Data security (0 / 1)						+
5. Ethical aspects (0 / 1)						+
6. Other (0 / 1)						+

Duomenų talpyklos pasirinkimas

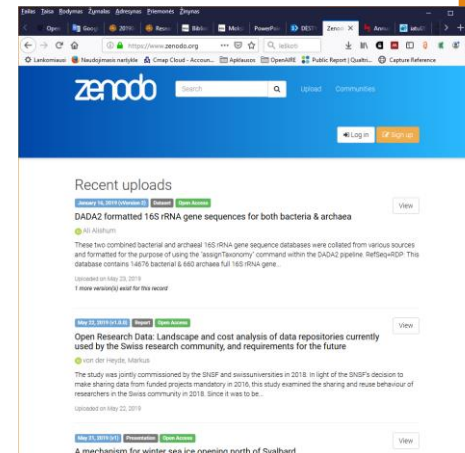
Pasirinkti teminę talpyklą, kurioje duomenys saugomi pagal atitinkamos disciplinos standartus



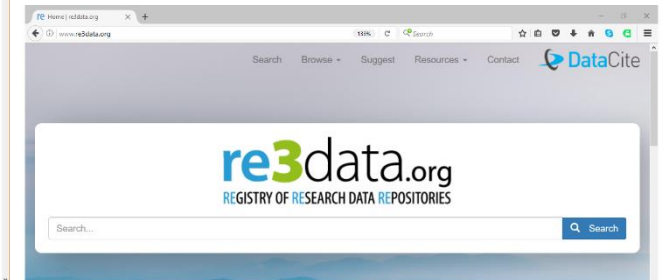
Institucinę mokslo duomenų talpyklą ar kitą institucinę duomenų saugyklą



Naudoti nemokamą talpyklą, pvz. Zenodo



Ieškoti duomenų talpyklos registre <http://re3data.org>

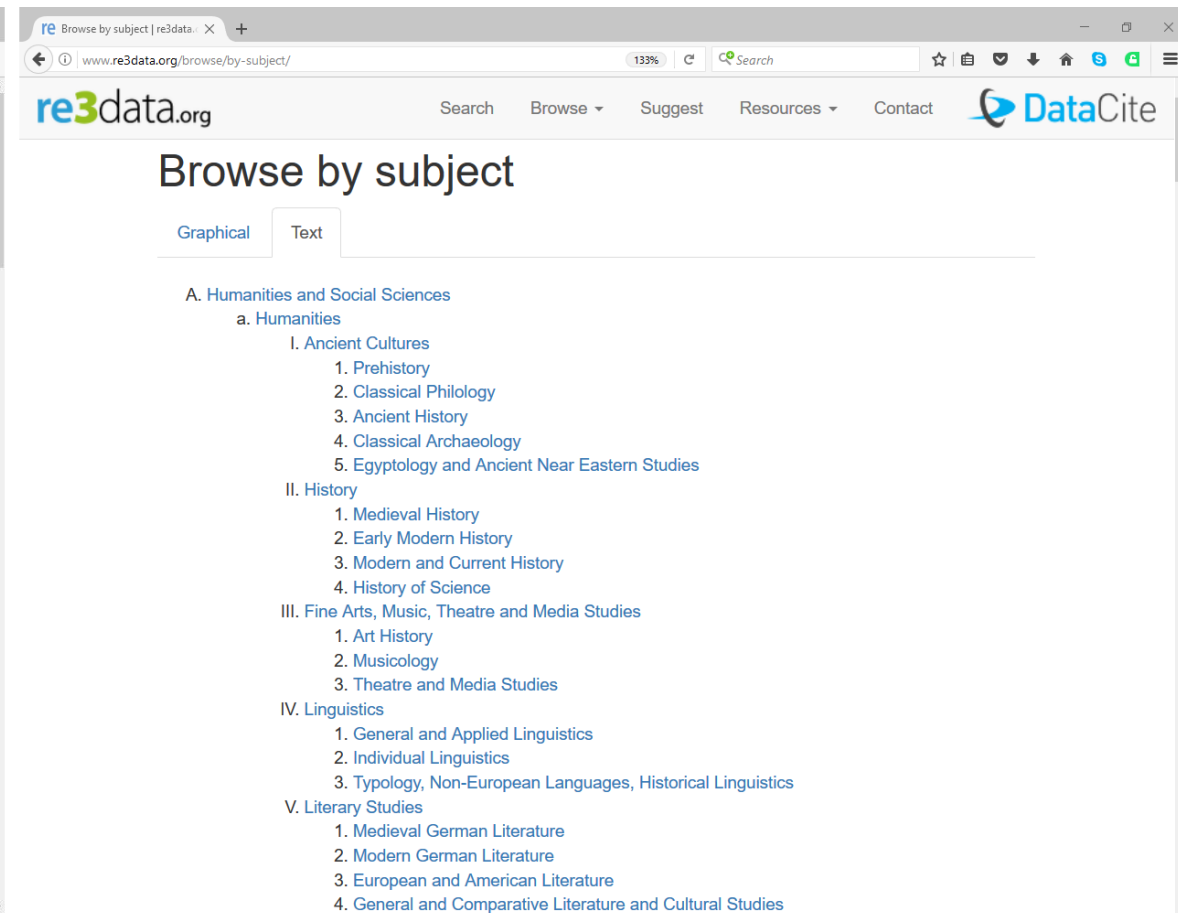
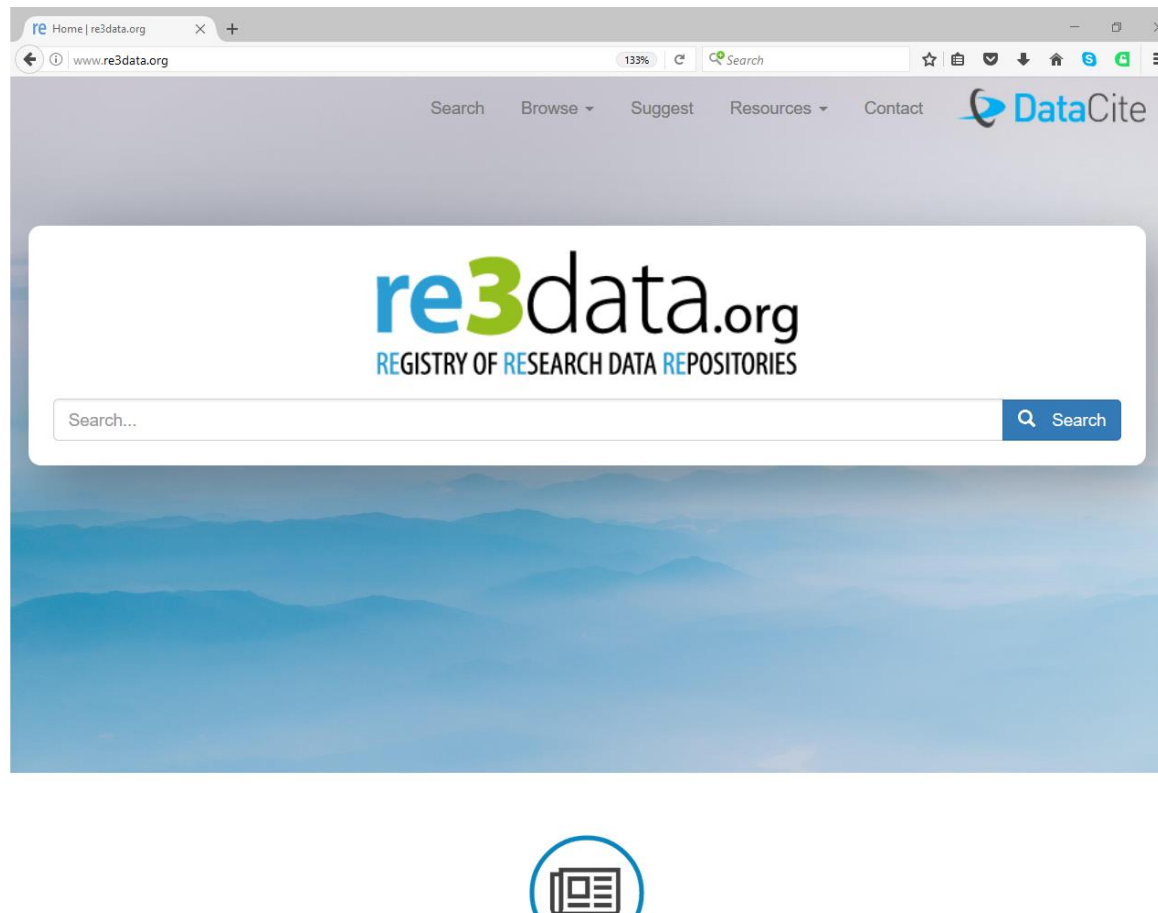


Daugiau informacijos: <https://www.openaire.eu/opendatapilot-repository>

Zenodo: <http://www.zenodo.org>

Re3data.org: <http://www.re3data.org>

Duomenų talpyklų registras re3data.org



Multidisciplininės duomenų talpyklos

- Dataverse (<https://dataverse.harvard.edu>)
- FigShare (<http://figshare.com>)
- Dryad (<https://datadryad.org>)
- Mendeley Data (<https://data.mendeley.com>)
- Zenodo (<http://zenodo.org>)
- DataHub (<http://datahub.io>)
- DANS (<http://www.dans.knaw.nl>)
- EUDat (<http://www.eudat.eu>)



Duomenų talpyklos pasirinkimo kriterijai



- Atitinka poreikius: tinkami duomenų saugojimo formatai; atvira, apribota prieiga arba prieiga pagal pareikalavimą;
- Suteikia duomenims nuolatinį identifikatorių, pvz. DOI;
- Kaupiami duomenys, juos aprašantys metaduomenys ir įrankiai, užtikrinantys duomenų atgaminimą
- Pateikia rekomendacijas, kaip cituoti įkeltus duomenis;
- Susieja duomenis su publikacijomis, autoriais, finansavimo šaltinių ar projektų duomenimis.

Data Seal of Approval: <http://datasealofapproval.org/en>

Nestor seal: http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html

ISO 16363: <http://www.iso16363.org>

Duomenų valdymo ir FAIR duomenų principų įgyvendinimas

- Ugdyti duomenų valdymo gebėjimus
- Rengti tyrėjams pagalbą teikiančius darbuotojus
- Duomenų tvarkymą ir citavimą įtraukti į mokslo vertinimo sistemą
- Skirti tikslinį finansavimą duomenų tvarkymui

Atvirosios prieigos prie mokslo publikacijų ir duomenų gairės

20. Projekto vadovas turi užtikrinti projekto įgyvendinimo metu gautų duomenų išsaugojimą skaitmeniniu formatu, o pasibaigus projektui – perdavimą saugoti institucijai ir (ar) pateikimą talpyklai. Duomenys turi būti išlaikyti ne trumpiau kaip 5 metus pasibaigus projektui.

21. Projekto, kurio metu bus kaupiami duomenys, vykdytojai paraiškoje pateikia duomenų valdymo planą. Duomenų valdymo planas projekto eigoje gali būti tikslinamas.

22. Duomenų valdymo planui įgyvendinti projekto laikotarpiu patiriamos išlaidos gali būti dengiamos Tarybos finansuojamų projektų lėšomis ir numatomos projekto išlaidų sąmatoje.

23. Duomenys, kurių pagrindu parengtos Gairių IV skyriuje nurodytos mokslo publikacijos, turi būti atvirai prieinami tuo pačiu metu, kaip ir publikacijos. Šie duomenys turi būti prieinami talpyklose ar publikacijų leidėjų nurodytais kitais būdais, ir susieti su atitinkamomis publikacijomis.

24. Projektų vykdytojams, pateikus motyvuotą paaiškinimą, Gairių nuostatos dėl atvirosios prieigos prie duomenų (ar jų dalies) gali būti netaikomos, jeigu:

- 24.1. duomenys nebuvo gauti projekto įgyvendinimo metu arba mokslo publikacija nėra paremta originaliais duomenimis, t. y. jie nebuvo surinkti ir (ar) sukurti projekto metu;
- 24.2. reikia apsaugoti rezultatus, ketinant duomenis panaudoti komercinimo ar pramonės vystymo tikslais;
- 24.3. duomenų atvėrimas nesuderinamas su konfidencialumo reikalavimais;
- 24.4. duomenų atvėrimas prieštarautų asmens duomenų apsaugos reikalavimams;
- 24.5. duomenų atvėrimas trukdytų pasiekti projekto tikslus;
- 24.6. yra kitų teisėtų priežasčių neatverti duomenų.

Skyriai	Klausimai	Paaškinimai
A. Duomenų rinkimas	1. Kokius duomenis planuojate rinkti ar sukurti?	1.1. Ar egzistuoja duomenys, kuriuos galėtumėte pakartotinai panaudoti? 1.2. Kokie numatomi duomenų tipai, formatas ir apimtys?
B. Duomenų ir jų atsarginių kopijų kaupimas	2. Kaip duomenys ir jų atsarginės kopijos bus kaupiami projekto metu?	2.1. Kur bus kaupiami duomenys? 2.2. Kokiu būdu planuojate atkurti duomenis, jei jie būtų pažeisti? Ar kursite duomenų atsargines kopijas?
	3. Kaip užtikrinsite sukauptų duomenų saugumą?	3.1. Kokia gali būti rizika duomenų saugumui ir kaip ši rizika bus valdoma? 3.2. Kaip užtikrinsite, kad projekto partneris (jei yra) galėtų saugiai pasiekti duomenis?
C. Duomenų atranka ir saugojimas	4. Kurie duomenys yra ilgalaikės vertės ir turi būti saugomi?	4.1. Kurie duomenys turi būti saugomi ar sunaikinami dėl tam tikrų sutartinių nuostatų, teisinių ar kitų reikalavimų? 4.2. Kokį laikotarpį duomenys bus saugomi?
D. Prieiga prie duomenų	5. Kaip užtikrinsite duomenų prieinamumą ir galimybę jais naudotis?	5.1. Kada duomenys taps prieinami? 5.2. Kaip potencialūs naudotojai sužinos apie duomenis? 5.3. Kam bus suteikta galimybė naudotis duomenimis ir kokiomis sąlygomis?
E. Atsakomybė ir ištekliai	6. Kam bus priskirta atsakomybė už duomenų tvarkymą bei valdymą?	6.1. Kas atsakingas už DVP įgyvendinimą bei periodišką peržiūrą ir koregavimą? 6.2. Ar nuostatos dėl duomenų nuosavybės ir atsakomybės už mokslinių tyrimų duomenų tvarkymą / valdymą bus aptartos su partneriu (jei yra)?
	7. Kokie žmogiškieji ir kiti ištekliai bus reikalingi rengiant bei įgyvendinant DVP?	7.1. Ar reikės įdarbinti specialias kvalifikacijas turintį darbuotoją? 7.2. Ar bus reikalinga speciali, papildoma įranga, įskaitant programinę? 7.3 Ar įvertinote, kad duomenų bankai ar saugyklos gali taikyti mokesčius už duomenų kaupimą, saugojimą ar atvėrimą?

(Projekto mokslinės ataskaitos forma)

5.4. Duomenų valdymo plano įgyvendinimas

..... projekto
(Nurodoma Tarybos remiamos veiklos kryptis)

Jei paraiškoje buvo planuojama kaupti duomenis, pateikiamas duomenų valdymo plano įgyvendinimo pagrindimas – aprašoma, kokie duomenys buvo sukaupti, kokie standartiniai metaduomenys buvo naudojami juos apibūdinant, nurodoma, kaip ir kokiomis sąlygomis jie yra (jei taikomas embargo laikotarpis – bus) prieinami pakartotiniam naudojimui, kada ir kaip jie bus atveriami kitiems vartotojams, kaip gauti duomenys bus tvarkomi ir atnaujinami; jei duomenys nėra (nebus) atverti, pateikiamas to pagrindimas, nurodomos teisinės ir (ar) techninės kliūtys, trukdančios atvirai prieigai prie duomenų.

Tolimesni žingsniai

- Rengiant DVP numatyti duomenis tvarkyti pagal FAIR principus;
- Parengti duomenų ir infrastruktūrų tvarkymo specialistus, ekspertus;
- Propaguoti naujus mokslinės komunikacijos būdus (duomenų leidyba, duomenų citavimas);
- Įtraukti duomenų tvarkymą ir citavimą į mokslo vertinimo sistemą;
- Įgalinti automatinę turinio analizę (data and text mining);
- Skatinti dalintis duomenimis ir juos pakartotinai panaudoti;
- Finansuoti duomenų rengimą, tvarkymą ir palaikymą (EK siūlymas: 5 proc. nuo projektui skiriamo finansavimo);
- Suderinti aukštojo mokslo politiką su ES reikalavimais.

Mokymosi kursai

- Efektyvus elektroninių mokslo informacijos išteklių naudojimas
<https://open.ktu.edu/course/view.php?id=37>
- Research Academy
<https://researcheracademy.elsevier.com/research-preparation/research-data-management>
- CESSDA training
<https://www.cessda.eu/Training>