

Internetinis seminaras: Atvirojo mokslo ir tyrimų duomenų aktualijos

***FAIR duomenys socialinių ir
humanitarinių mokslų
atstovams***



2020-06-01

Dr. Vaidas Morkevičius ir Doc. Rimantas Rauleckas
Kauno technologijos universitetas
Lietuvos HSM duomenų archyvas (LiDA)

Planas

Duomenų valdymas ir atviri duomenys

Rimantas Rauleckas

Klausimai ir atsakymai

Kaip atverti savo duomenis laikantis FAIR duomenų principų

Vaidas Morkevičius

Klausimai ir atsakymai

Atviri duomenys/mokslas – strateginis prioreitetas (H2020 etc.)

- Atverti reikia ne tik publikacijas, bet ir duomenis (programinę įrangą ir kitą mokslinę produkciją)
- Gėrybės sukurtos už viešus pinigus turi būti viešos

Duomenų valdymo planai (reikalauja daugmaž visos finansuojančios institucijos)

- Kaip bus valdomi projektų duomenys, kur jie bus padėti ir tvarkomi? Ar bus vieši (jei įmanoma)?

Esminis duomenų atvėrimo principas:

- As Open as Possible, as Closed as Necessary
- Duomenų atvėrimą galima pagrįstai apriboti:
 - Trumpam – kol projekto vykdytojai parengs savo publikacijas ir produkciją
 - Ilgam – siekiant užtikrinti duomenų subjektų konfidencialumą

Ar pakanka, kad duomenys būtų atviri?

- Duomenų **atvėrimo kokybė**? (Open Access vs. Open/FAIR data)
- 2014 m. surengtas seminaras/dirbtuvės „Jointly Designing a Data Fairport“ Leidene (Olandija)
 - The meeting concluded with a draft formulation of a set of foundational principles that were subsequently elaborated in greater detail – namely, that all **research objects should be Findable, Accessible, Interoperable and Reusable (FAIR) both for machines and for people.** These are now referred to as the **FAIR Guiding Principles.** (Wilkinson et al.)



FAIR duomenų principai

Tik dar vienas standartas, ar kažkas tikro/ilgalaikio?

- Atrodo, **ilgalaikis dalykas**

Ką jis reiškia?

- **Kaip įgyvendinti** (mokslininkams, mokslo administratoriams etc.)
- **Kaip išmatuoti** „FAIRness“?

Research data will not become nor stay FAIR by magic. We need skilled **people**, **transparent processes**, interoperable **technologies** and collaboration to build, operate and maintain **research data infrastructures**.

Mari Kleemola, Finnish Social Science Data Archive/CoreTrustSeal Board, Secretary

<https://tietoarkistoblogi.blogspot.com/2018/11/being-trustworthy-and-fair.html>

Images by kenwoodpress.com, Good Ware by flaticon.com, freebeesupply.com, openlibrary.org

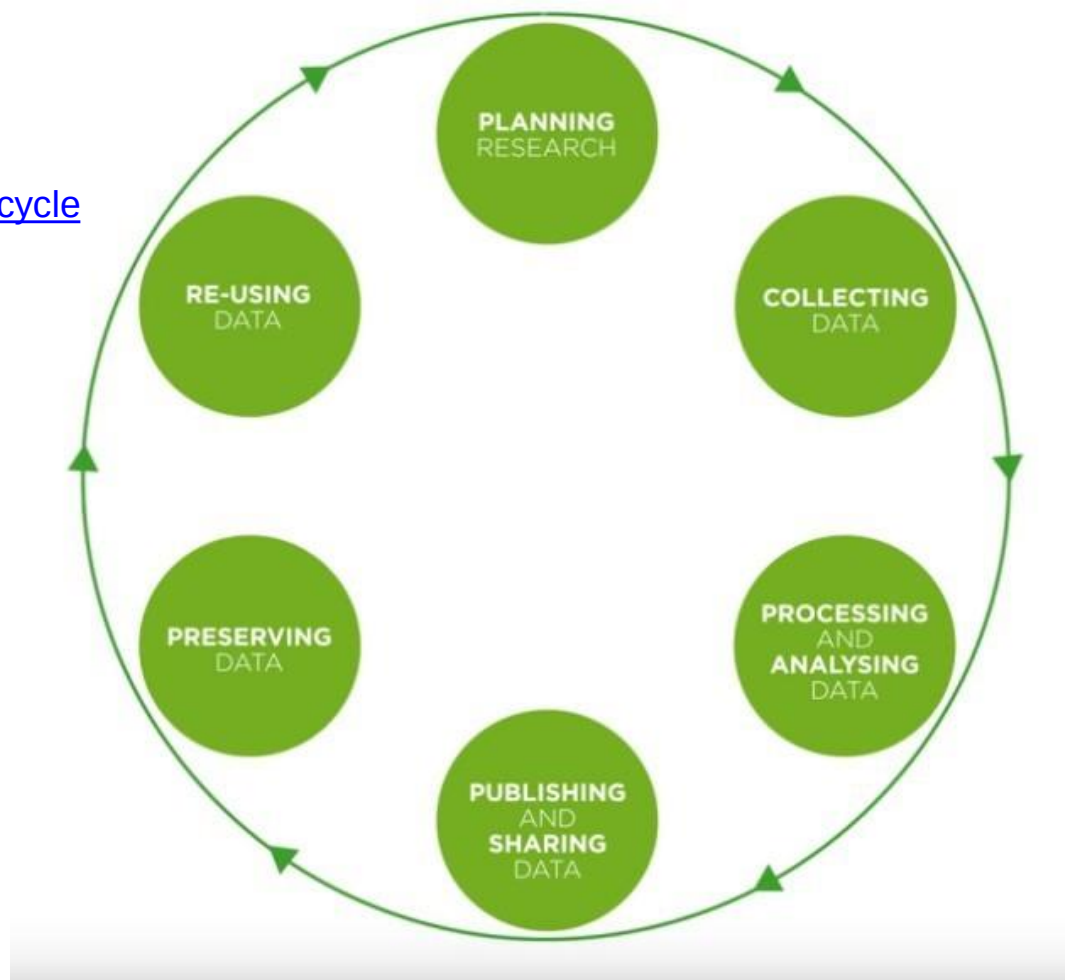


FAIR duomenų principai

Duomenų gyvavimo ciklas

UK Data Archive:

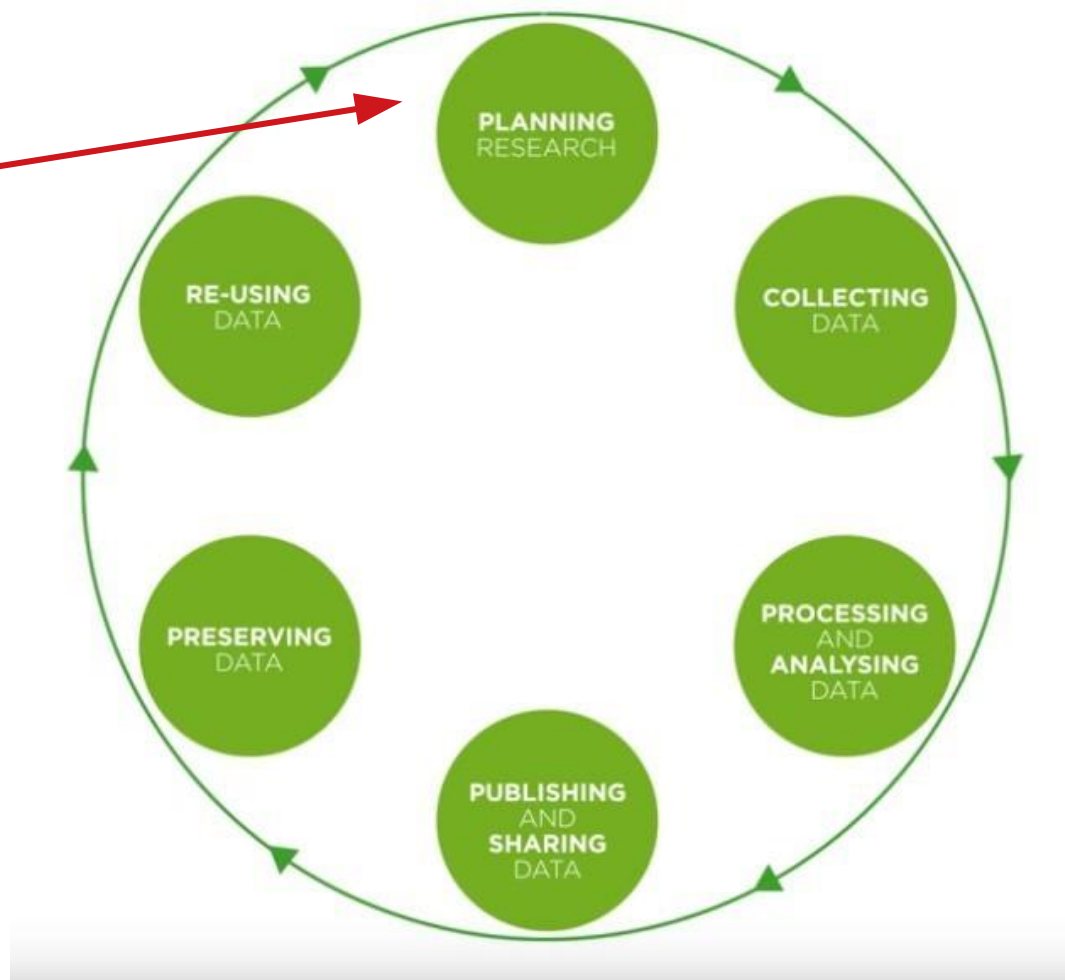
www.ukdataservice.ac.uk/manage-data/lifecycle



FAIR duomenų principai

Duomenų gyvavimo ciklas

Paprastai mokslininkas pradeda
nuo mokslinio tyrimo planavimo



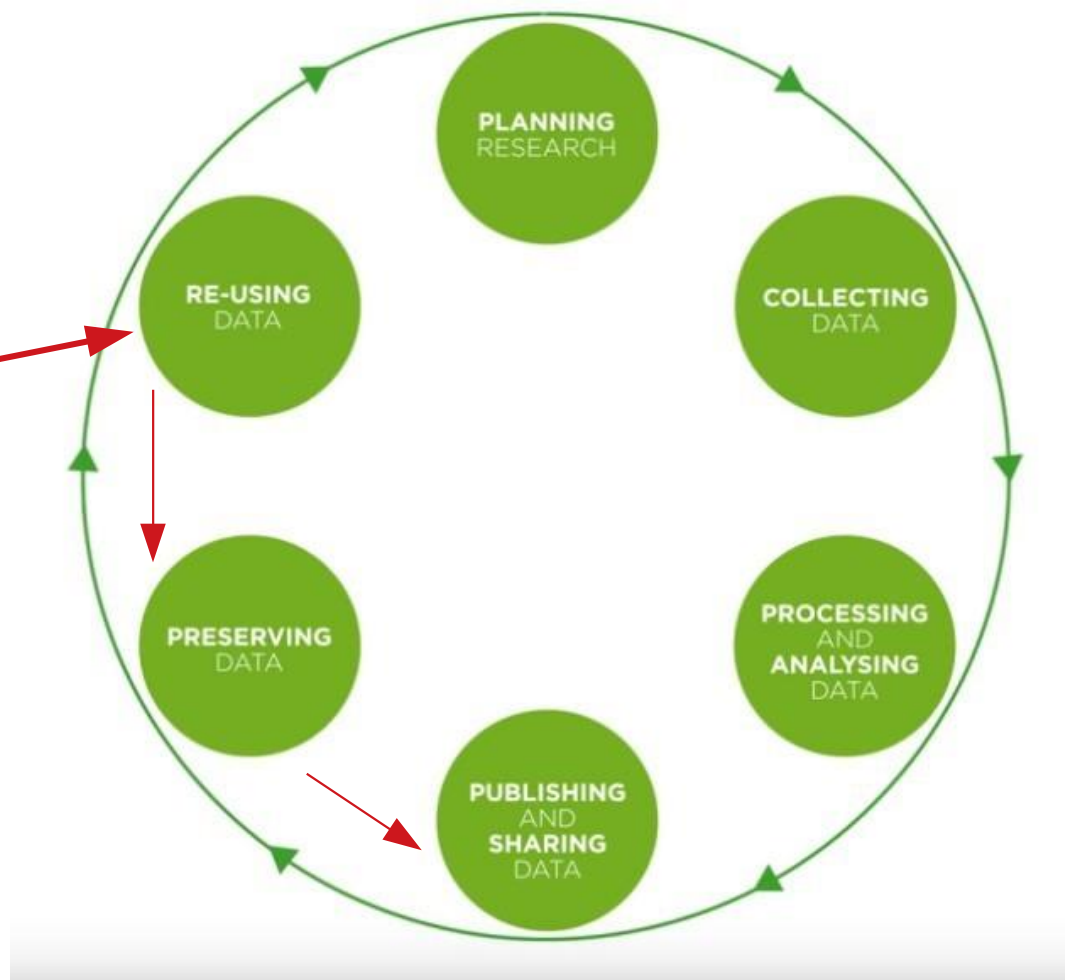
FAIR duomenų principai

Duomenų gyvavimo ciklas

Paprastai mokslininkas pradeda nuo mokslinio tyrimo planavimo

Tačiau kalbant apie duomenų atvėrimą **logika atvirkštinė**

- Pakartotinis duomenų naudojimas
- Reikia žinoti/išmanyti duomenis, kad galėtum juos tinkamai panaudoti



FAIR duomenų principai

Patikimumas (pasitikėjimas duomenimis) yra svarbiausia

- Žmogus pakartotinai naudojantis duomenis nori žinoti:
 - Iš kur atsirado šie duomenys?
 - Kaip jie buvo sukurti?
 - Kaip jie buvo apdorojami?
- Duomenis sukūręs žmogus nori būti tikras, kad:
 - Sukurti duomenys bus tinkamai naudojami ir interpretuojami

FAIR duomenų principai nurodo **gaires, kaip** skaitmeninius resursus (pvz., duomenų rinkinius) **padaryti randamais, pasiekiamais, sąveikiais ir pakartotinai panaudojamais** viso duomenų gyvavimo ciklo metu

www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

https://vidensportal.deic.dk/sites/default/files/uploads/A%20FAIRy%20tale%20book%20digital_ny.pdf

FAIR duomenų principai

www.go-fair.org/fair-principles, www.force11.org/group/fairgroup/fairprinciples

Randami (Findable)

- Svarbiausias dalykas – rasti duomenis, todėl metaduomenis ir duomenis turėtų lengvai rasti tiek žmonės, tiek kompiuteriai. Kompiuteriu perskaitomi metaduomenys įgalina automatinius duomenų ir paslaugų radybos algoritmus.
 - F1. (Meta)duomenims priskirtas globaliai unikalus ir nuolatinis identifikatorius (PID → DOI)
 - F2. Duomenys aprašyti turtingais metaduomenimis (žr. apibrėžimus prie R1)
 - F3. Į metaduomenis aiškiai ir konkrečiai įtraukti duomenų, kuriuos jie apibūdina, identifikatoriai
 - F4. (Meta)duomenys registruoti ar indeksuojami talpyklose su paieškos galimybėmis

FAIR duomenų principai

www.go-fair.org/fair-principles, www.force11.org/group/fairgroup/fairprinciples

Pasiekiami (Accessible)

- Kai randami reikiami duomenys, būtina žinoti, kaip juos galima pasiekti, galimai naudojant tinkamą autentifikaciją ar autorizaciją.
 - A1. (Meta)duomenys yra prieinami pagal identifikatorius naudojant standartinius komunikacijos protokolus
 - A1.1 Protokolai yra atviri, nemokami ir universaliai įgyvendinami (HTTP, FTP, telefonas)
 - A1.2 Protokolai leidžia naudoti tinkamas autentifikacijos ir autorizacijos procedūras, kai reikia (FAIR ≠ OPEN)
 - A2. Metaduomenys išlieka prieinami, net jei duomenys tampa neprieinamais

FAIR duomenų principai

www.go-fair.org/fair-principles, www.force11.org/group/fairgroup/fairprinciples

Sąveikūs (Interoperable)

- Duomenys paprastai turi būti integruojami su kitais duomenimis. Be to, duomenys turi sąveikauti su jų apdorojimo, saugojimo ir analizės programomis.
 - I1. (Meta)duomenyse informacijai reprezentuoti naudojama formali, prieinama, bendra ir plačiai taikoma kalba (XML, RDF)
 - I2. (Meta)duomenyse naudojami žodynai, kurie suderinti su FAIR principais (CESSDA <https://vocabularies.cessda.eu/#!discover>)
 - I3. (Meta)duomenyse naudojamos reikiamos ir tinkamos nuorodos į kitus (meta)duomenis

FAIR duomenų principai

www.go-fair.org/fair-principles, www.force11.org/group/fairgroup/fairprinciples

Pakartotinai panaudojami (Reusable)

- Esminis FAIR principų tikslas – optimalus pakartotinis duomenų naudojimas. Šiam tikslui pasiekti duomenys ir metaduomenys turi būti gerai aprašyti, kad juos būtų galima atkartoti ar sujungti su kitais duomenimis kituose kontekstuose.
 - R1. Meta(duomenys) yra turtingai aprašyti daugeliu tikslų ir svarbių elementų ar požymių
 - R1.1. (Meta)duomenys išleisti su aiškia ir prieinama duomenų naudojimo licencija (<https://creativecommons.org/licenses>)
 - R1.2. (Meta)duomenys susieti su detalia kilmės/sukūrimo informacija (autorystė)
 - R1.3. (Meta)duomenys atitinka (mokslo) kryptiškai svarbius bendruomenės nustatytus standartus
 - <https://rd-alliance.github.io/metadata-directory/standards>
 - www.dcc.ac.uk/guidance/standards/metadata
 - <https://fairsharing.org>, https://tiny.cc/DANS_File_Formats

FAIR duomenų principai

Kaip pamatuoti duomenų „FAIRness“?

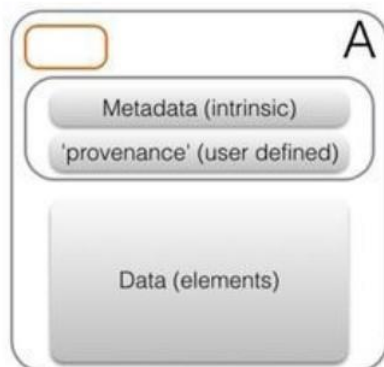
(<https://github.com/FAIRMetrics/Metrics>)

- EUDAT bendras sąrašas („check list“) <https://zenodo.org/record/1065991>
 - Pabandykim kartu nustatyti šio duomenų rinkinio FAIR statusą:
 - <https://doi.org/10.7910/DVN/L9YNV6>
- Australijos mokslinių tyrimų bendruomenės (ARDC) „matuoklis“ <https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool>
 - Pabandykim kartu nustatyti šio duomenų rinkinio FAIR statusą:
 - www.lidata.eu/data/quant/LIDA_ESS_0308
- Olandų archyvo (DANS) „matuoklis“ <https://satisfyd.dans.knaw.nl>
 - Pabandykim kartu nustatyti šio duomenų rinkinio FAIR statusą:
 - <http://hdl.handle.net/20.500.11821/17>

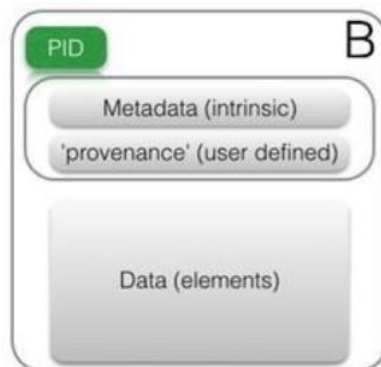
FAIR duomenų principai

Data as increasingly FAIR Digital Objects

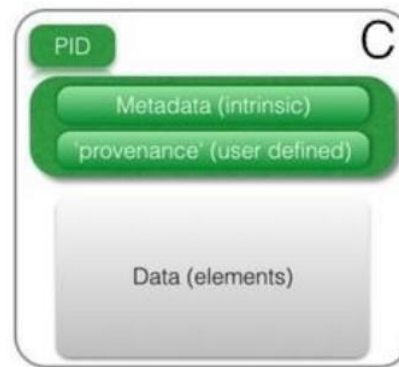
Re-useless data (80%)



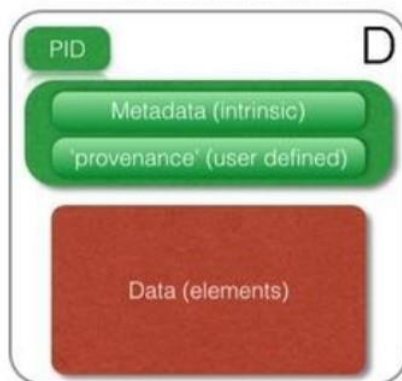
Findable



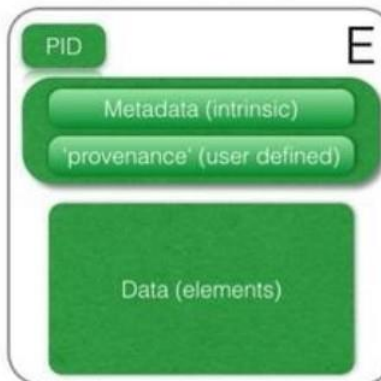
FAIR metadata



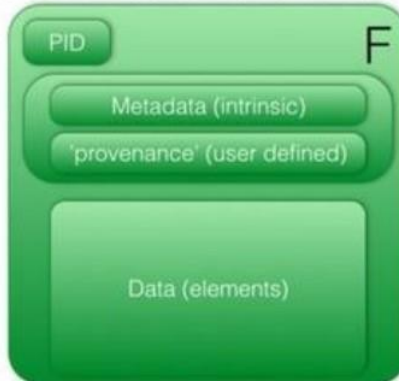
FAIR data-
restricted access



FAIR data-
Open Access



FAIR data-
Open Access/Functionally Linked



O gal paprasčiau patikimos **duomenų talpyklos**?

Duomenų talpyklos yra virtualios saugyklos, kuriose saugomi, tvarkomi ir teikiama prieiga prie įvairių formatų skaitmeninių resursų

- Institucinės (visos įstaigos ar padalinio)
- Mokslo krypčių ar šakų
- Bendro pobūdžio

Visos gali turėti tam tikrų **reikalavimų** dėl

- Pakartotinio naudojimo
- Failų formatų ir duomenų struktūros
- Naudojamų metaduomenų standartų

Kodėl vartotojams (mokslininkams) turėtų rūpėti FAIR duomenys, duomenų talpyklos ir su FAIR principais suderinamos duomenų talpyklos?

O gal paprasčiau patikimos **duomenų talpyklos**?

Duomenų **talpyklų privalumai**:

- Palengvina gyvenimą mokslininkams (mažiau rūpesčių dėl tinkamo duomenų atvėrimo)
- Stiprina mokslinį sąžiningumą ir pasitikėjimą
- Sukurti duomenys tampa
 - Prieinamais
 - Suprantamais
 - Pakartotinai panaudojamais
- **Duomenų talpyklos jūsų duomenis padaro FAIR ir saugo pagal FAIR principus**

O gal paprasčiau patikimos **duomenų talpyklos**?

Kokiu būdu talpyklose **duomenys** tampa **FAIR**?

Pavyzdžiui:

- Suteikiamas nuolatinis unikalus **identifikatorius** (PID)
 - Užtikrinamas ilgalaikis randamumas ir tinkamas citavimas
- Užtikrinamas randamumas per **viešus paieškos katalogus**
 - Efektyvus duomenų radimas yra esminė dalijimosi duomenimis prielaida
- Užtikrinamas tinkamas duomenų ir metaduomenų **licencijavimas**
 - Aiškūs terminai ir sąlygos
- Įgyvendinami ir skleidžiami **(meta)duomenų standartai**
 - Sąveikumas

O gal paprasčiau patikimos **duomenų talpyklos**?

Kaip duomenys talpyklose **saugomi pagal FAIR**?

- Domenų talpyklos užtikrina **ilgalaikį** FAIR skaitmeninių objektų **kuravimą**, todėl garantuoja, kad duomenys laikui bėgant išliks FAIR
 - Konsultuoja ir remia **duomenų kūrėjus** (pvz., (meta)duomenų standartai)
 - Konsultuoja ir remia **duomenų vartotojus** (pvz., citavimas)

O gal paprasčiau patikimos **duomenų talpyklos**?

Kaip **surasti** tinkamą **duomenų talpyklą**?

- Pasirinkti **sertifikuotą** talpyklą
- Naudoti **mokslo šakos ar krypties** talpyklą
- Naudoti **institucinę** talpyklą (jei užtikrina ilgalaikį saugojimą)
- Bendros talpyklos – tik jei nėra aukščiau išvardytų
- Talpyklų **paieška**:
 - <https://repositoryfinder.datacite.org>
 - www.re3data.org
 - <https://ddrs-dev.dariah.eu/ddrs>

O gal paprasčiau patikimos **duomenų talpyklos**?

Patikimos talpyklos – sertifikuotos

- **Reguliariai vertinamos** pagal nustatytus kriterijus
- **Teikia tam tikras paslaugas**, kurios užtikrina duomenų rinkinių suderinamumą su FAIR principais
- Egzistuoja keli sertifikavimo būdai, kuriais galima remtis vertinant duomenų talpyklos patikimumą (ISO, DIN)
 - Labiausiai paplitusi sertifikavimo sistema: **CoreTrustSeal**
www.coretrustseal.org
- Svarbu ir IT infrastruktūra
 - **Dataverse** bene geriausiai suderinama su FAIR principais: www.dataverse.org



O gal paprasčiau patikimos **duomenų talpyklos**?

- **Patikimumas** – svarbiausias dalykas dalijantis ir pakartotinai naudojant duomenis
- **FAIR principų matavimas gali būti vykdomas įvairiais lygiais** ir turi apimti infrastruktūrą (talpyklą)
- Su FAIR principais suderinamos duomenų talpyklos užtikrina **ilgalaikį duomenų prieinamumą, suprantamumą ir pakartotinį naudojimą**
- **FAIR duomenys „gyvena“ patikimose duomenų talpyklose**
 - Sertifikuotas talpyklos FAIR duomenis saugo pagal FAIR principus
 - CoreTrustSeal reikalavimai suderinami su FAIR duomenų principais

O gal paprasčiau patikimos duomenų talpyklos?

CORETRUSTSEAL – FAIR ALIGNMENT

Mokrane, 2018

