



www.data.europa.eu

Metadata Quality Assurance

04.04.2026

Metadatan laadunvarmistusmetodologia

How data.europa.eu measures the quality of harvested metadata

The Metadata Quality Assessment (MQA) is a tool developed by the consortium of data.europa.eu to study the quality of metadata harvested by data.europa.eu. It is intended to help data providers and data portals to check their metadata quality and to receive suggestions for improvements. The results are presented via the MQA and are also available as download. In the following we describe the functionality of the MQA and the methodology it uses.

If this page still does not answer all your questions, please feel free to contact us via our feedback form at the end of the page.

[Metadatan laadunvarmistusmetodologia](#)

Tutkimuksen laajuus

Metadatan laadunvarmistuksen avulla halutaan vastata seuraavaan kysymykseen:
Mikä on metadatan laatu julkisen sektorin datalle yleiseurooppalaisella alueella ja mitkä ovat suurimmat esteet datan laadun parantamiselle?

Tämän perusteella metadatan laadunvarmistus tutkii seuraavia konkreettisia kysymyksiä:

- Yhteensopivuus DCAT-AP:n ja DCAT-AP-johdannaisten kanssa
- Sellaisten tietojen luovuttaminen, joille DCAT-AP ei ole pakollinen
- Metadatatassa viitatus datan saavutettavuus verkko- ja latausosoitteissa
- Viitatus datan koneellinen luettavuus
- Lisenssien käyttö

Jokaisesta kysymyksestä seuraa yksittäisiä tutkimuksia, joita on kuvattu tarkemmin jäljempänä.

Mitä ei käsitellä

The MQA is limited by the metadata it can examine. The investigation is limited exclusively to the metadata that data.europa.eu collects during the harvesting process. If there are errors in the source metadata, these can falsify the overall result. To limit this error potential, the MQA provides a validation service that can be used by data providers to validate their metadata for valid formats and compliant DCAT-AP before integrating it into the harvesting process.

[DCAT-AP SHACL validation service web page](#)

Metadatan laadunvarmistusprosessi

Metadatan laadunvarmistus tarkistaa metadatan kunkin haravoinnin yhteydessä. Metadatan laadunvarmistus mittaa eri kriteerien laatua. Kriteerit on selostettu alla olevissa taulukoissa. Tarkistusten tulokset tallennetaan datan laatusanastona ([DQV](#)). DQV on W3C:n spesifikaatio, jota käytetään tietoaaineiston laadun kuvailemiseen.

Saavutettavuus voi vaihdella, joten verkko- ja latausosoitteet on tarkistettava toistuvasti. Siksi metadatan laadunvarmistus tarkistaa kaikkien jakelujen saavutettavuuden säännöllisesti. Muiden kriteerien tarkistuksesta poiketen tämä tarkistus kestää kauemmin, sillä jakelut tarkistetaan HTTP:n kautta, ja kullakin pyydetyllä verkko-osoitteella voi olla pitempi vasteaika. Metadatan laadunvarmistuksen käyttämä mekanismi huomioi sen, että jokaisen verkko-osoitteen saavutettavuus tarkistetaan uudelleen muutaman viikon kuluessa edellisestä tarkistuksesta.

Olettamukset

Metadatan laadunvarmistus perustuu seuraaviin oletuksiin.

Ei-pakollisten kenttien käyttö

Uskomme, ettei pelkkä pakollisten DCAT-AP-kenttien käyttö riitä laadukkaan metadatan tarjoamiseen. Siksi arviointi tarkistaa myös kentät, joita ei ole määritetty pakollisiksi DCAT-AP:n mukaisesti. Tarkistettavat kentät on lueteltu alla.

Sama sisältö useissa jakeluissa

Jos tietoaaineisto sisältää useamman kuin yhden jakelun, kaikki jakelut ovat sisällöltään samanlaisia ja eroavat vain datan esitystavan osalta. Tietoaaineistolla voi olla esimerkiksi kaksi jakelua, joista toinen tarjoaa datan PDF-muodossa ja toinen saman datan koneellisesti luettavassa RDF/XML-muodossa.

Dimensiot

Tässä osiossa on kuvattu kaikkia dimensioita, joita metadatan laadunvarmistus tutkii laatua määritellessään. Dimensiot on johdettu [FAIR-periaatteista](#).

Löydettävyys

Seuraavassa taulukossa on kuvattu mittareita, jotka auttavat ihmisiä ja koneita tietoaisteistojen etsinnässä. Tällä alueella voi saada enintään 100 pistettä.

kriteeri	Kuvaus	Mittari	Paino
Avainsanojen käyttö	Avainsanat tukevat hakua suoraan ja parantavat siten tietoaisteiston löydettävyttä.	Järjestelmä tarkistaa, onko avainsanoja määritetty. Avainsanojen määrä ei vaikuta pistemäärään. Dataset dcat:keyword	30
Luokat	Luokat auttavat käyttäjiä tutkimaan tietoaisteistoja aihealueen mukaan.	Tarkistetaan, onko tietoaisteistolle määritetty yksi vai useampi luokka. Luokkien määrä ei vaikuta pistemäärään. Dataset dcat:theme	30
Maantieteellinen haku	Paikkatietojen avulla käyttäjät voivat etsiä tietoaisteistoa maantieteellisellä haulla.	Tarkistetaan, onko ominaisuus määritetty vai ei. Dataset dcat:spatial	20
Aikaperusteinen haku	Väli aikaisten tietojen avulla käyttäjät voivat tehdä aikaperusteisia hakuja.	Tarkistetaan, onko ominaisuus määritetty vai ei.	20

Saavutettavuus

Seuraavassa taulukossa on kuvattu mittareita, joiden avulla määritetään, onko jakelujen viittaama data varmasti saavutettavissa. Tällä alueella voi saada enintään 100 pistettä.

kriteeri	Kuvaus	Mittari	Paino
Verkko-osoitteen saavutettavuus	Verkko-osoite ei ole välttämättä suora linkki dataan, vaan se voi myös viitata verkko-osoitteeseen, jonka kautta tietoaimeistoon pääsee tai jossa on siitä lisätietoja.	Määritetyn verkko-osoitteen saavutettavuus tarkistetaan HTTP HEAD -pyynnöllä. Jos vastaukseksi saatu tilakoodi on 200- tai 300-alueella, resurssin saavutettavuus arvioidaan positiivisesti. Distribution dcat:accessURL	50
Latausosoite	Latausosoite on suora linkki viitattuun dataan.	Tarkistetaan, onko ominaisuus määritetty vai ei. Distribution dcat:downloadURL	20
Latausosoitteen saavutettavuus	Jos latausosoite on annettu, saavutettavuus tarkistetaan.	Määritetyn verkko-osoitteen saavutettavuus tarkistetaan HTTP HEAD -pyynnöllä. Jos vastaukseksi saatu tilakoodi on 200- tai 300-alueella, resurssin saavutettavuus arvioidaan positiivisesti. Distribution dcat:downloadURL	30

Yhteentoimivuus

Seuraavassa taulukossa on kuvattu mittareita, joilla määritellään, voidaanko jakelu katsoa yhteentoimivaksi. Sama sisältö, useita jakeluja -olettamuksen mukaisesti pisteiden laskennassa käytetään vain jakelua, jolla on suurin pistemäärä. Tällä alueella voi saada enintään 110 pistettä.

kriteeri	Kuvaus	Mittari	Paino
Formaatti	Tässä kentässä määritetään jakelun tiedostomuoto.	Tarkistetaan, onko ominaisuus määritetty vai ei. Distribution dct:format	50
Mediatyyppi	Tässä kentässä määritetään jakelun mediatyyppi.	Tarkistetaan, onko ominaisuus määritetty vai ei. Distribution dcat:mediaType	10
Formaatti/mediatyyppi sanastosta	Tarkistaa, kuuluvatko formaatti ja mediatyyppi valvottuun sanastoon.	The format vocabulary can be found in the data.europa.eu GitLab repository . The media type is check against the IANA list Distribution dct:format dcat:mediaType	10
Ei-yksinoikeudellinen	Tarkistaa, onko jakelun formaatti ei-yksinoikeudellinen.	The distribution is considered as non-proprietary if the specified format and media type is contained in the corresponding data.europa.eu GitLab repository vocabulary. Distribution dct:format	20
Koneellisesti luettava	Tarkistaa, onko jakelun formaatti koneellisesti luettava.	The distribution is considered as machine-readable if the specified format and media type is contained in the corresponding data.europa.eu GitLab repository vocabulary. Distribution dct:format	20
DCAT-AP-yhteensopivuus	DCAT-AP compliance is calculated across all sources and datasets available on a catalogue. This check is only performed if the metadata is originally harvested as DCAT-AP or as a valid derivate. DCAT-AP is a specification for describing linked public data in Europe. The data.europa.eu portal may also harvest metadata which does not fully comply to DCAT-AP. In order to increase conformity to DCAT-AP, the MQA checks each metadata for its DCAT-AP compliance.	The metadata is validated against a set of SHACL shapes . The metadata is not compliant, if the SHACL validation reports at least one issue. The MQA uses data.europa.eu's DCAT-AP SHACL validation service . SHACL is a recommendation from the W3C and is used for validating RDF graphs against a set of shapes.	30

Uudelleenkäytettävyys

Seuraavassa taulukossa on kuvattu, mitä mittareita käytetään datan uudelleenkäytettävyyden tarkistamiseen. Tällä alueella voi saada enintään 75 pistettä.

kriteeri	Kuvaus	Mittari	Paino
Lisenssitiedot	Lisenssi on arvokasta tietoa datan uudelleenkäyttöä ajatellen.	Tarkistetaan, onko ominaisuus määritetty vai ei. Distribution dct:license	20
Lisenssisanasto	Virheellisistä lisenssitiedoista kertovia ilmoituksia halutaan rajoittaa. Usein törmätään esimerkiksi CC-lisensseihin, joista puuttuvat versiotiedot.	This section describes all dimensions that the MQA examines in order to determine the quality. The dimensions are derived from the <u>FAIR principles</u> . The MQA recommends and credits the usage of controlled vocabularies. The data.europa.eu portal publishes its controlled vocabularies in GitLab. The vocabularies are derived from the <u>EU Vocabularies</u> . Distribution dct:license	10
Käyttörajoitukset	Tämä kenttä ilmaisee, onko datan käyttöoikeus yleinen vai rajattu.	Tarkistetaan, onko ominaisuus määritetty vai ei. Dataset dct:accessRights dcat:mediaType	10
Käyttörajoitusten sanasto	Valvotun sanaston käyttö parantaa uudelleenkäytettävyttä.	Tarkistetaan, onko käyttöoikeuksien valvottu sanasto käytössä. Dataset dct:accessRights	5
Yhteyspiste	Yhteyspiste sisältää tietoja siitä, kenelle dataa koskevat kyselyt voi osoittaa.	Tarkistetaan, onko ominaisuus määritetty vai ei. Dataset dct:contactPoint	20
Julkaisija	Julkaisija on datan julkaissut henkilö tai organisaatio.	Tarkistetaan, onko ominaisuus määritetty vai ei. Dataset dct:publisher	10

Tilannekohtaisuus

Seuraavassa taulukossa on joitakin kevyitä ominaisuuksia, jotka tarjoavat lisää kontekstia käyttäjälle. Tällä alueella voi saada enintään 20 pistettä.

kriteeri	Kuvaus	Mittari	Paino
Oikeudet	Joissakin tapauksissa tietoaaineistoon ei voi käyttää tiettyä lisenssiä. Oikeudetkentässä voi määrittää viittauksen resurssiin, jossa kerrotaan käyttäjälle hänen oikeuksistaan tietoaaineistoa käytettäessä.	Tarkistetaan, onko ominaisuus määritetty vai ei. Distribution dct:rights	5
Tiedostokoko	Määrittää tiedoston koon tavuina.	Tarkistetaan, onko ominaisuus määritetty vai ei. Distribution dct:byteSize	5
Julkaisupäivä	Päivämäärä, jona tietoaaineisto tai jakelu julkaistiin.	Tarkistetaan, onko ominaisuus määritetty vai ei. Dataset and Distribution dct:issued	5
Muokkaustiedot	Päivämäärä, jona tietoaaineistoa tai jakelua muutettiin viimeksi.	Tarkistetaan, onko ominaisuus määritetty vai ei. Dataset and Distribution dct:modified	5

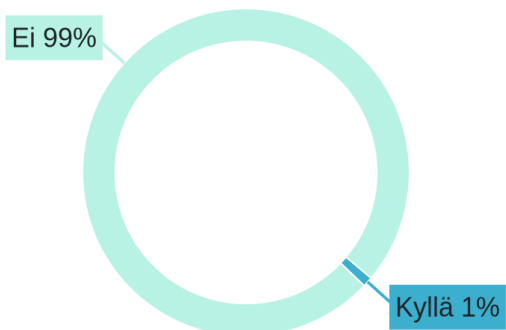
Luokitus

Lopullinen luokitus tapahtuu neljän luokitusryhmän kautta. Pisteiden yhdistäminen luokitusluokkaan on esitetty alla olevassa taulukossa. Luokituksen edustus metadatan laadunvarmistuksessa ilmaistaan yksinomaan luokitusluokkien avulla. Näin datan tuottajat voivat saavuttaa korkeimman luokituksen, vaikka pisteitä hieman vähennettäisiin.

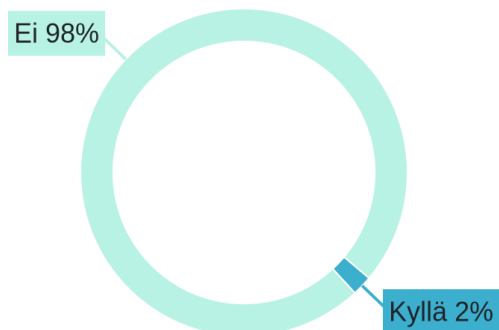
Dimensio	Enimmäispistemäärä
Löydettävyys	100
Saavutettavuus	100
Yhteentoimivuus	110
Uudelleenkäytettävyys	75
Tilannekohtaisuus	20
Summa	405

Luokitus	Pistemääräväli
Erinomainen	351 - 405
Hyvä	221 - 350
Kohtalainen	121 - 220
Huono	0 - 120

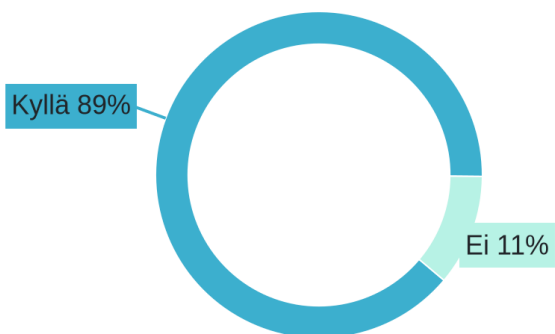
Oikeudet



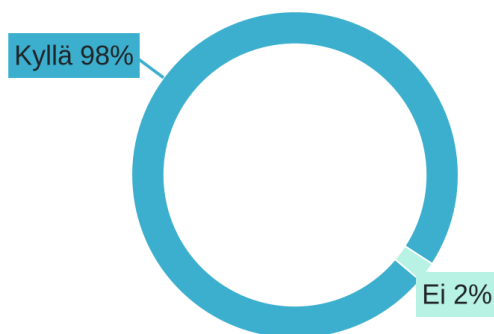
Tiedostokoko



Julkaisupäivä

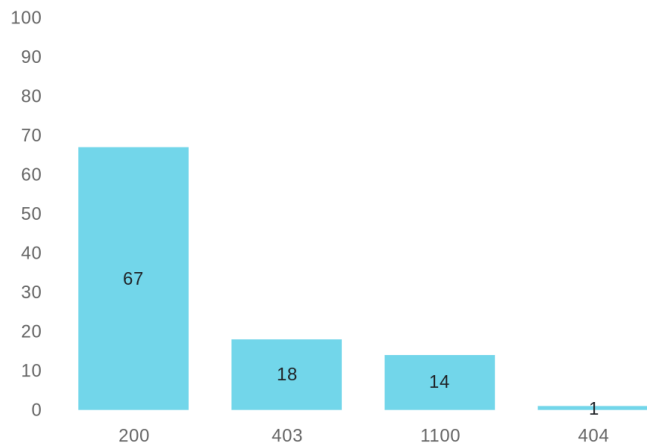


Muokkaustiedot

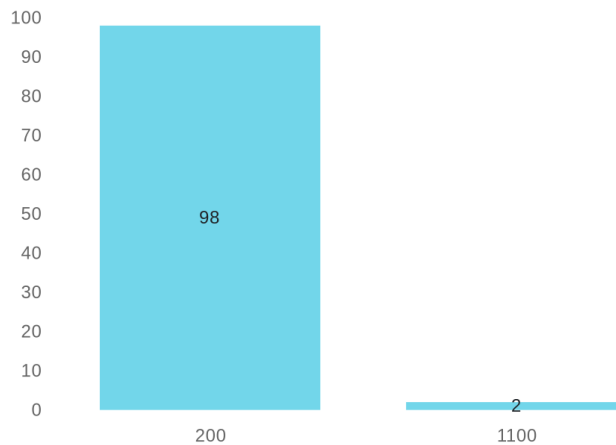


Saavutettavuus

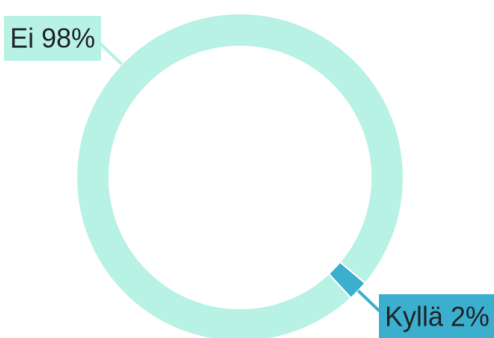
Useimmin käytettyjen verkko-osoitteiden tilakoodit



Useimmin käytettyjen latausosoitteiden tilakoodit

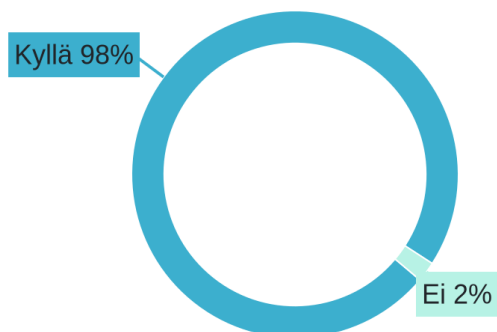


Latausosoite

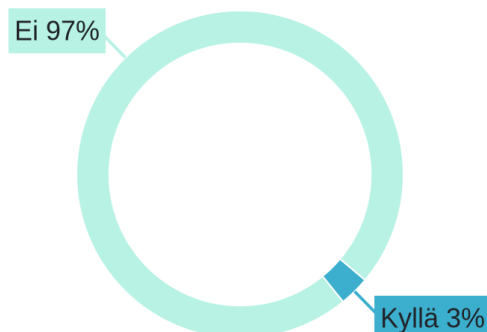


Uudelleenkäytettävyys

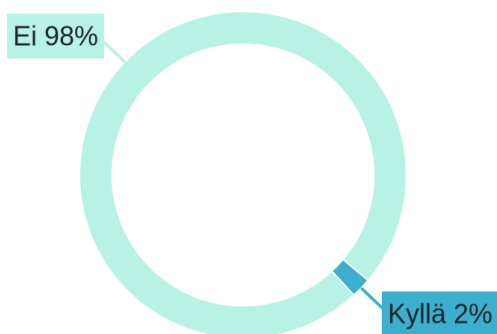
Lisenssitiedot



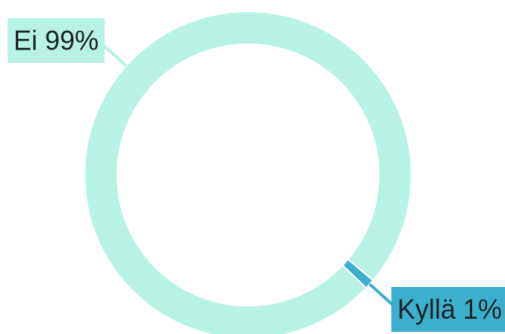
Lisenssisanasto



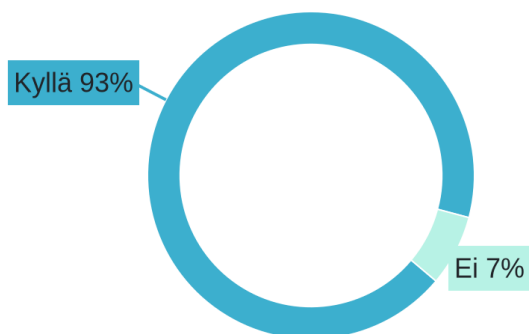
Käyttörajoitukset



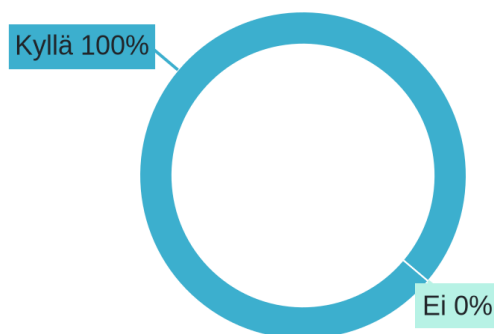
Käyttörajoitusten sanasto



Yhteyspiste

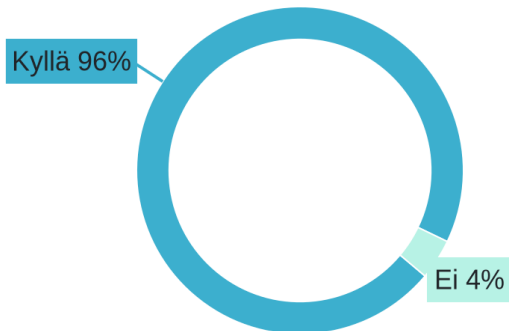


Julkaisija

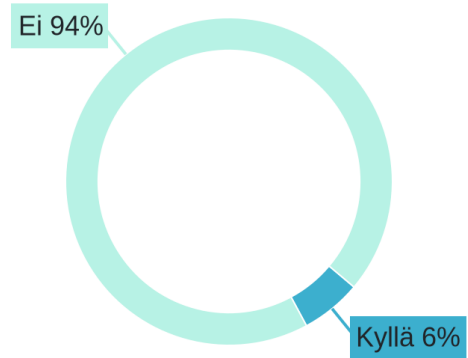


Yhteentoimivuuus

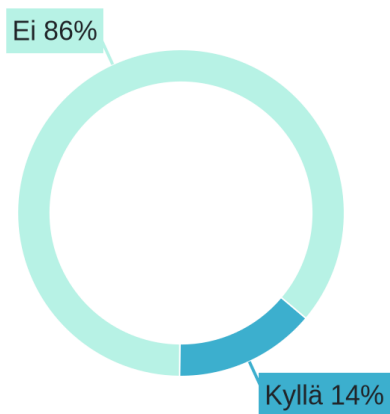
Muoto



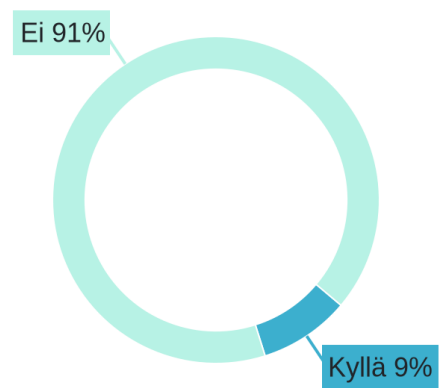
Mediatyyppi



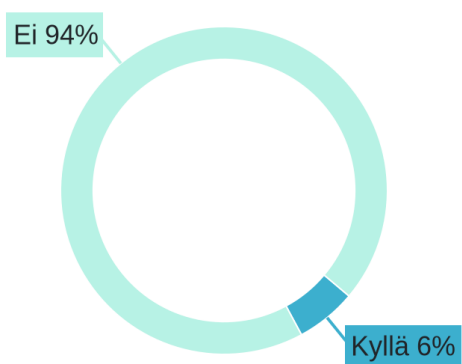
Formaatti/mediatyyppi sanastosta



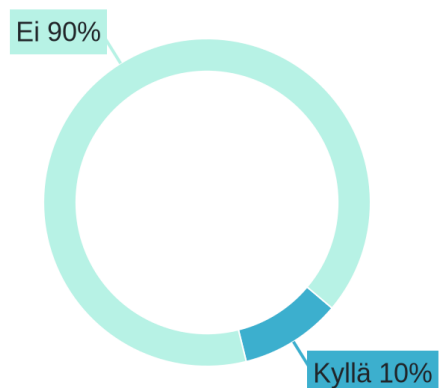
Ei-yksinoikeudellinen



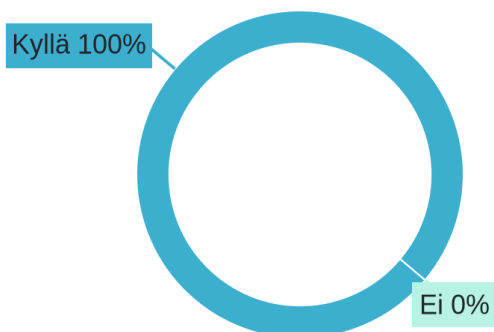
Koneellisesti luettava



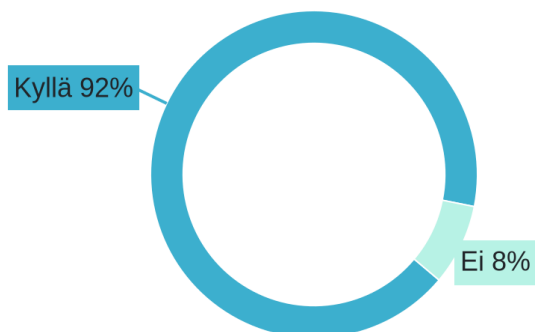
DCAT-AP-yhteensopivuuus



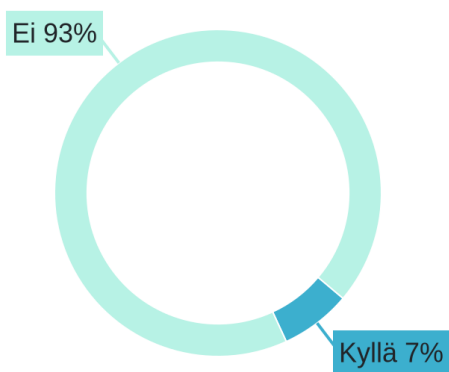
Avainsanojen käyttö



Luokat



Maantieteellinen haku



Aikaperusteinen haku

