



# Uniform metadata for Finnish repositories

Tanja Vienonen, National Library of Finland  
Open Repositories 2017, 30.6.2017

# Some background



Image credit:  
Riitta Koikkalainen

- Tajua - Tieteen avoin julkaiseminen, or the Open Scientific Publishing project
- A national project to promote open access principles at Finnish libraries and improve the availability of Finnish research

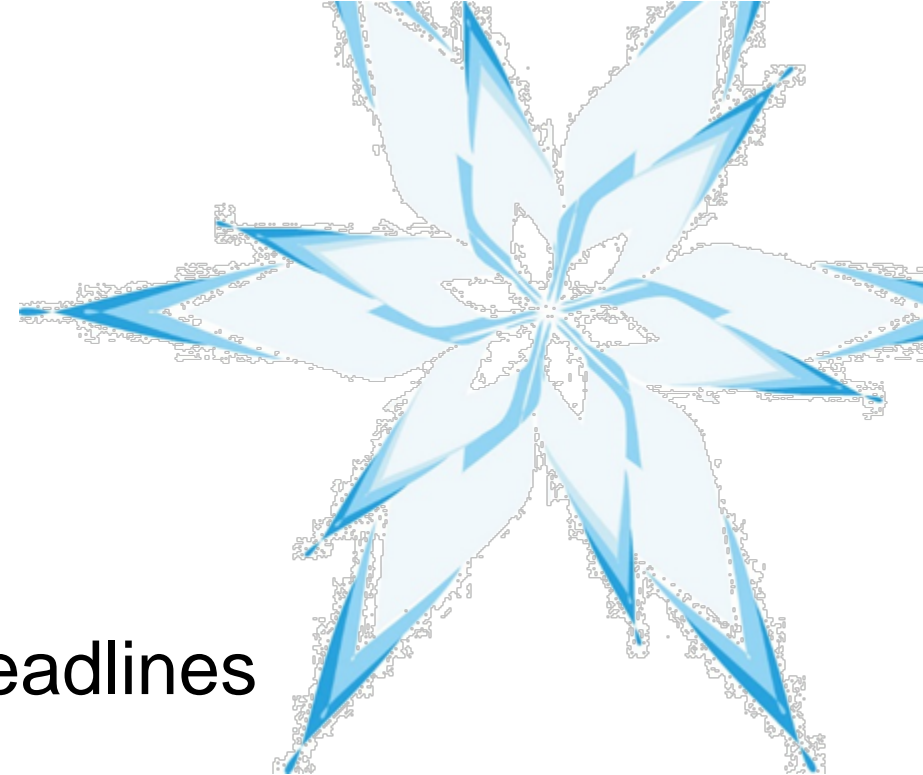
# A bit more

- In Finland, more than 60 organisations have an institutional repository
- 13 repository instances in total, 7 of them hosted by NatLib
- All except one based on DSpace
- Some are shared repository instances of up to 27 organisations
- All use Qualified Dublin Core
- Over 300,000 open access full-text publications



# Initial challenges

- Heterogeneous metadata practices
- Ad hoc solutions made in a rush to meet deadlines
- Non-standard Dublin Core fields
- Aggregation of repository metadata problematic

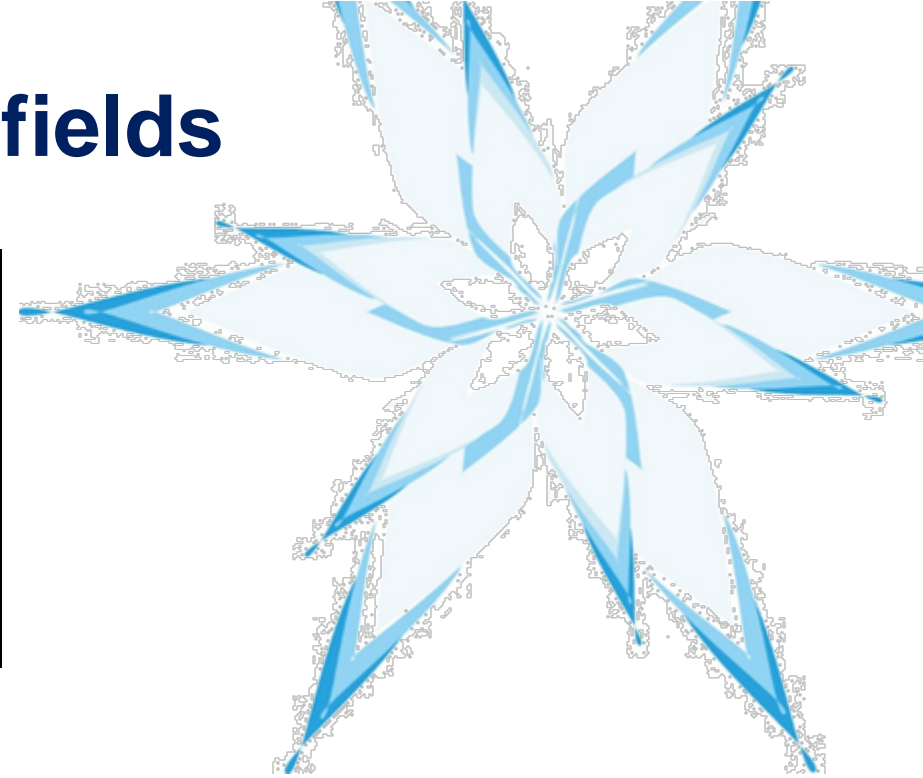


# Examples of non-standard metadata fields

dc.kansainvalisyys  
dc.kansallinenyhteisjulkaisu.muu  
dc.kansallinenyhteisjulkaisu.vsti  
dc.kansallinenyhteisjulkaisu.yosp  
dc.kansallinenyhteisjulkaisu  
dc.keyword.other  
dc.kvyhteisjulkaisu

dc.series.issn  
dc.series.volume  
dc.series  
dc.sizebyperiods

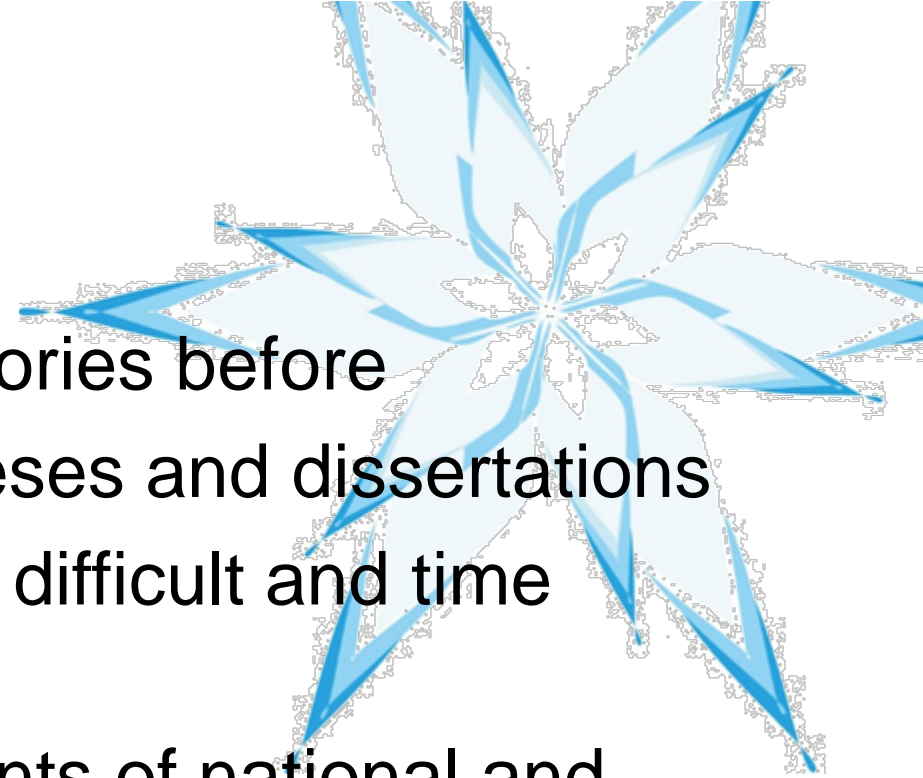
dc.evojulkaisu  
dc.fileformat.other  
dc.fileformat



Source: doria.fi

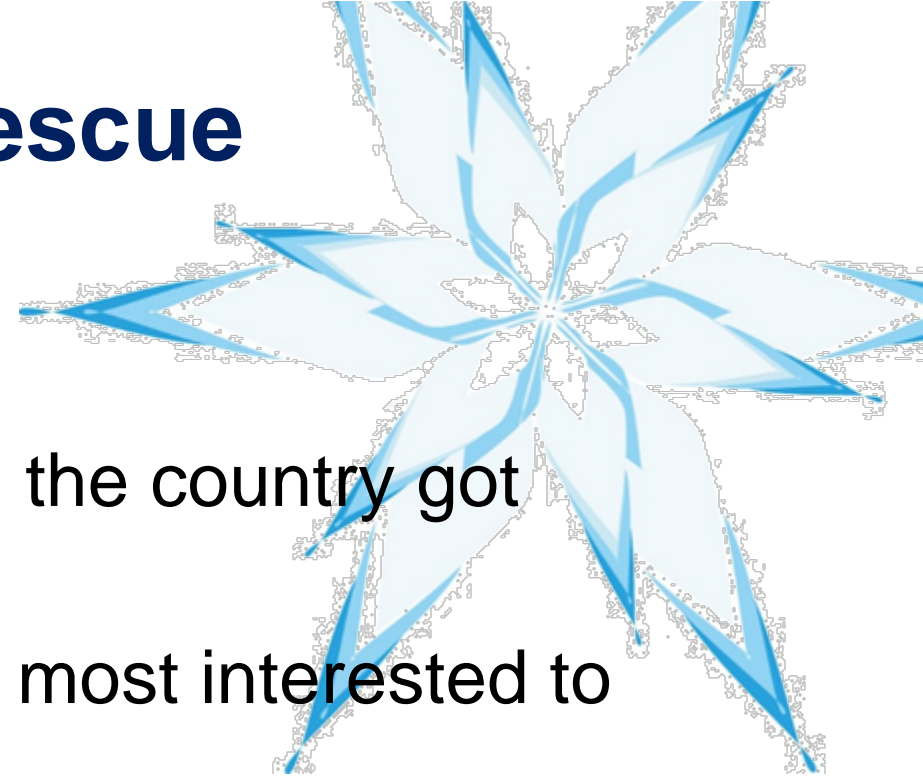
# Repository managers in distress

- No general metadata guidelines for repositories before
- Partly outdated metadata guidelines for theses and dissertations
- Maintenance, development and migrations difficult and time consuming
- Difficulties in complying with the requirements of national and international metadata aggregators
- Repository managers had to fend for themselves



# Metadata working group to the rescue

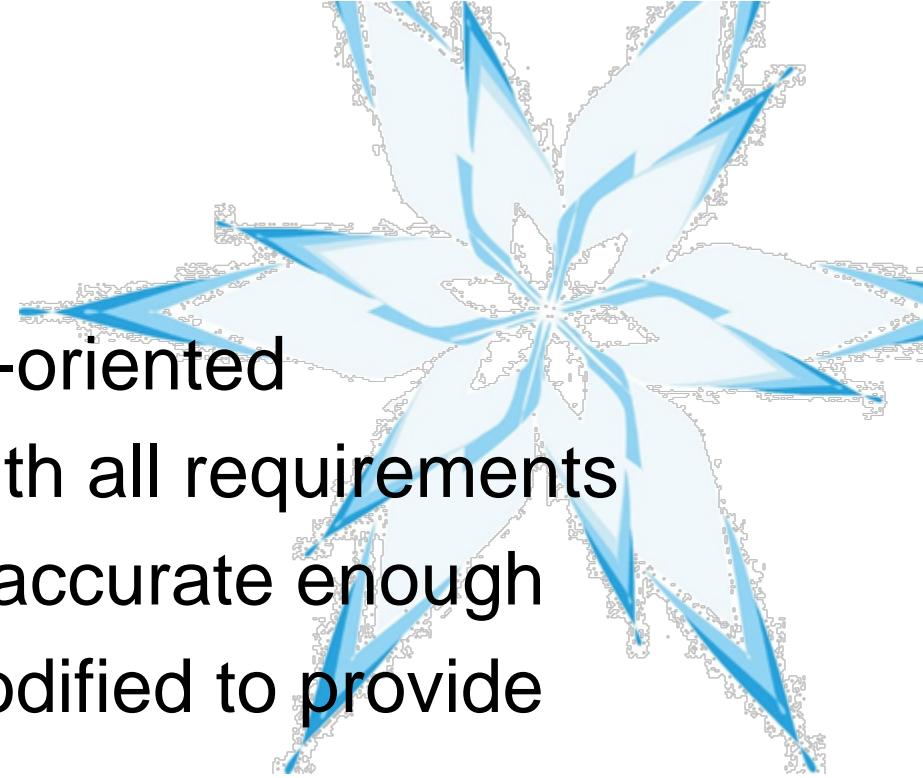
- Over 80 experts from organisations around the country got involved in the project
- Smaller working group formed from people most interested to contribute
- Several NatLib experts actively involved in the work and a few more were consulted on specific aspects of metadata





# Priorities

- Approach to the project was practical, goal-oriented
- A single metadata format cannot comply with all requirements
- Aim to develop an internal format rich and accurate enough
- Repository crosswalks and APIs can be modified to provide metadata that fits aggregators' needs
- Semantic uniformity prioritised over having exactly “the right” metadata fields





# Desperately seeking metadata practices

- Exploring common metadata formats, standards, and aggregators
- Compiling and comparing most used metadata fields in the repository metadata
- Suggesting metadata fields closer to standard Dublin Core as basis for general guidelines



# Snippet from the 1<sup>st</sup> draft

[Hahmotelma] Julkaisuarkistojen tekstiaineiston metadatasuosituks

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

€ % 0.00 123 Arial 8 B I A

Metadatakenttä

	A	B	C	D	E	F	G
1	Metadatakenttä	Dublin Core	Lisätietoa/tallennusmuoto	Pakollinen	Toistettava	Ristiriitaisia käytötapoja	Huomioita
18	Kontrolloitu asiasanasto	dc.subject [asiasanaston lyhenne] Esim. dc.subject.yso	Kontrolloidusta sanastosta tarkistetut asiasanat. Yksi asiasana per kenttä, ei ketjutettuja asiasanoja. *Soveltuu pakollinen: OpenAIRE *Finna *OKM: tieteenalat (pakollinen tieto)?		X	AaltoDoc: dc.subject.other	Sana-alojen lyhenteet ensisijaisesti LC:n lisäalta: <a href="http://www.loc.gov/standards/subjectlist/subject.html">http://www.loc.gov/standards/subjectlist/subject.html</a>
19	Tiivistelmä	dc.description.abstract	Vapaatekstikenttä julkaisun tiivistelmälle/abstraktille. Kielikoodien käyttö suositeltavaa. Suositukset abstraktin kopioimiseen: UTF-8 mukaisia. *Finna		X		Tähtäykseksi erotella erikelliset abstraktit jotenkin? Ja tunnistaako lomake esim. kappaleajoat? Tekninen kysymyskuin formaatin asä. Erikeliset abstraktit problemaattinen erityisesti jos niitä on monta. UTF-8 mukaisia, koska XML-koodi menee rikki eikä tietuetta voi enää lukea. Pitäisikö silti olla haravoitavissa, vaikka täällä hetkellä on mainittu vain Finnan vaatimuksissa?
20	Sisällysluettelo	dc.description.tableofcontents	Monografian(?) sisällysluettelo *Finna				
21	Huomautukset	dc.description.notification	Vapaatekstikenttä erilaisille huomautuksille, käytössä esim. Theseuksessa, TamPubissa ja Valtossa. Julkinen kenttä. Lähdetietokannan lyhenne, jos metadata on tuotu jostain toisesta tietokannasta, pitää vastata identifier-kentässä annettua lähdetietokannan tunnistetta. Esim. WOS.		X	JYX: dc.description	On tarpeellinen.
22	Metadatan lähdetietokanta	dc.source.metadata					Tällä hetkellä sisäinen kenttä, voinee käyttää myös muista viittaustietokannoista tuotujen metadatojen merkintään.
23	Enemmän pohdintaa vaativat kentät						
24	Organisaatiotieto	dc.organization dc.provenance.organization? dc.relation.organization dc.relation.affiliation? dc.contributor.department?	Julkaisuarkistoa ylläpitävä organisaatio. dc.organization on järjestelmän sisäinen epästandardi kenttä, ei tule suositukseen? *OKM: organisaation alayksikkö (pakollinen: koodi)?				dc.organization ei virallinen DC-kenttä. Nimiavaruus olisi hyvä määrittellä. HUOM: katso myös molemmissa kentissä olevat kommentit
25	Muut tekijät	dc.contributor?	Kuvittajat, valokuvaajat, väitöskirjan ohjaajat, kustokset yms. *Suositeltu: OpenAIRE *Horizon 2020: E-U-rahoitustunnus ("kontrolloitu" temistö) *Finna: organisaatiotekijä ilman tarkennetta		X		Listattava muut. Julkaisun organisaatiotekijä, jos ei listata varsinaisia ihmistekijöitä. <a href="http://id.loc.gov/vocabulary/relatorshtml">http://id.loc.gov/vocabulary/relatorshtml</a> (LoC linked data relators) Opponentti: opn, väitöskirjan tarkastaja: rev, väitöskirjan ohjaaja: supervisor(?)
26	Artikkelin affiliaatiotieto	dc.contributor.department? dc.relation.affiliation? dc.relation.organization?	Tallennettavan artikkelin affiliaatio, eli missä tutkimuslaitoksessa se on tehty. *OKM: organisaation tekijät		X	JYX: dc.contributor.laitos dc.contributor.oppiaine	Tämä ei ratkaise tekijän ja oikean affiliaation yhdistämistä toisiinsa, mutta kertoisi kuitenkin missä tutkimus on tehty. DCMI: It is suggested that this affiliation be captured as the value of a dc:contributor property, recognising that the author's institution will have contributed to the creation of the resource. Because affiliation is a property of the resource, rather than a creator, the fact that there is no way to correlate particular authors with their affiliations (when there are multiple authors) is not an issue. ( <a href="http://dublincore.org/documents/dc-citation-guidelines">http://dublincore.org/documents/dc-citation-guidelines</a> )
27	Opinnäytteen tekopaikka	dc.contributor? dc.contributor.department?	*OKM?			Helda: yo, tdkja oppiaine erikseen	Opinnäytteen hyväksynyt vai tutkinnon hyväksynyt organisaatio? Tutkinnon hyväksynyt korkeakoulu ja laitos josta valmistunut?
28	Sivumäärä	dc.format.extent dc.items.SizeOrDuration	Sivujen kokonaislukumäärä julkaisuissa. *Suositeltu: OpenAIRE *Finna?			JYX ja DCMI: dc.format.extent Julkari: dc:journal.volumenumber	OpenAire: dc.format: <a href="https://guidelines.openaire.eu/en/latest/literature/field_format.html">https://guidelines.openaire.eu/en/latest/literature/field_format.html</a> MIME-tyyppi(?) käytössä. <a href="http://www.iana.org/assignments/media-types/media-types.xhtml">http://www.iana.org/assignments/media-types/media-types.xhtml</a> ppt: application/vnd.ms-powerpoint word: application/msword pdf: application/pdf

9 Lopullinen versio OpenAire Finna OKM MARC DC elementit CC-lisenssit 4 Hahmotelma

Explore

# Communication is key

- Spreadsheets collected into Google Drive for easy access and editing
- Meetings held online and in person
- Emails and NatLib's public online forum used to communicate developments



# Suddenly, it was published

## JULKAISUARKISTOPALVELUT



- Palvelun kuvaus
- Ajankohtaista
- Palveluun liittyminen
- ▼ Ohjeet
  - › Julkaisuarkistokohtaiset ohjeet
  - › Julkaisuarkiston käyttöliittymä ja työkalut
- ▼ Metadattaa koskevat ohjeet
  - › Julkaisuarkistokohtaiset metadatat
- ▼ Metadatasuositus julkaisuarkistojen tekstiaineistolle
  - ▼ 1 Kaikki metadatakentät
    - [dc.contributor.\[muu tekijyys\]](#)
    - [dc.contributor.\[tunnisteen lyhenne\]](#)
    - [dc.contributor.author](#)
    - [dc.contributor.editor](#)
    - [dc.contributor.organization](#)
    - [dc.creator](#)
    - [dc.date.available](#)
    - [dc.date.issued](#)

Pages > ... > [Metadatasuositus julkaisuarkistojen tekstiaineistolle](#)



## 1 Kaikki metadatakentät

Created by Tanja Vienonen, last modified on Apr 25, 2017

Alla on aakkosjärjestyksessä kaikki metadatasuosituksen Dublin Core -kentät tarkenteineen. Huomioi kirjainkoko kenttien nimityksissä erityisesti DSpace-julkaisuarkistoissa: kirjoita DC-kentät kokonaan pienillä kirjaimilla. Yksittäisen kentän otsikkoa klikkaamalla löytyy lisätietoa eri aggregoijista, esimerkkejä kentän sisällöstä ja muita huomioita kentän käyttötarkoituksesta.

Samat tiedot excel-taulukkona (esikatselun sijaan lataa omalle koneelle): [Metadatasuositus\\_julkaisuarkistojen\\_tekstiaineistolle\\_1.0.xlsx](#)

### [dc.contributor.\[muu tekijyys\]](#)

#### Muu tekijä

Vaihtoehtoja tarkenteeksi: degreeSupervisor, thesisAdvisor, opponent, artist, illustrator, photographer, translator, transcriber, other

Teoksen muut tekijät, kuten väitöskirjaohjaajat, vastaväittäjät, kuvittajat ja valokuvaajat. Jokainen nimi omaan kenttäänsä muodossa Sukunimi, Etunimi. Katso esimerkit [dc.contributor.author](#)-kentästä. Valitse tarkenne Library of Congressin "Code list for Relators"-sivustolta: <http://www.loc.gov/marc/relators/relaterm.html>. Toistettava kenttä.

### [dc.contributor.\[tunnisteen lyhenne\]](#)

#### Teokseen liittyvien henkilöiden tunnisteet

Kentän nimi esimerkiksi dc.contributor.orcid. Tekijän henkilökohtainen ORCID-tunniste tai muu tunniste, kuten ISNI. Jokainen tunniste omaan kenttäänsä. Toistettava kenttä.

# Nuts and bolts

- Final version was published in National Library's public wiki page
- Settled on 62 Dublin Core -based metadata fields with various qualifiers
- 11 extra metadata fields in a separate namespace in case repository is used for reporting the research publication metadata to the Ministry of Education and Culture's annual data collection



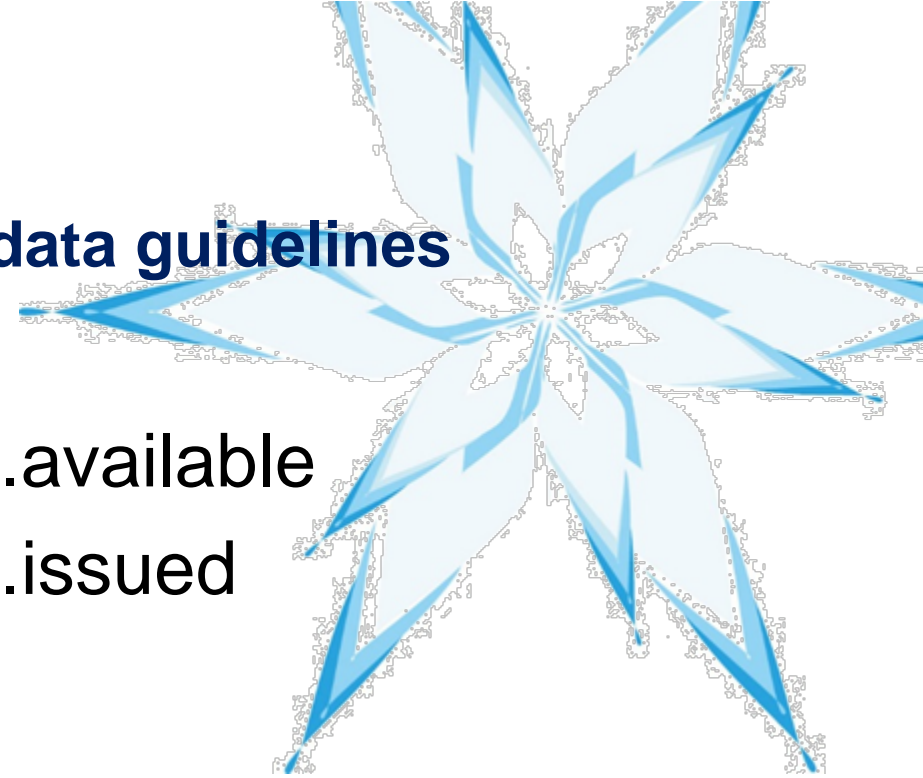
# Example: before and after

## Previous dc.date qualifiers

Item Status	Item Bitstreams	Item Metadata
<b>Add new metadata</b>		
<b>Name:</b>		
<div>dc.accrualmethod</div> <div>dc.date.accepted</div> <div>dc.date.accessioned</div> <div>dc.date.available</div> <div>dc.date.copyright</div> <div>dc.date.created</div> <div>dc.date.dateaccepted</div> <div>dc.date.datecontentscreated</div> <div>dc.date.earliestdatecreated</div> <div>dc.date.issued</div> <div>dc.date.latestdatecreated</div> <div>dc.date.submitted</div> <div>dc.date</div> <div>dc.description.abstract</div>		

## New metadata guidelines

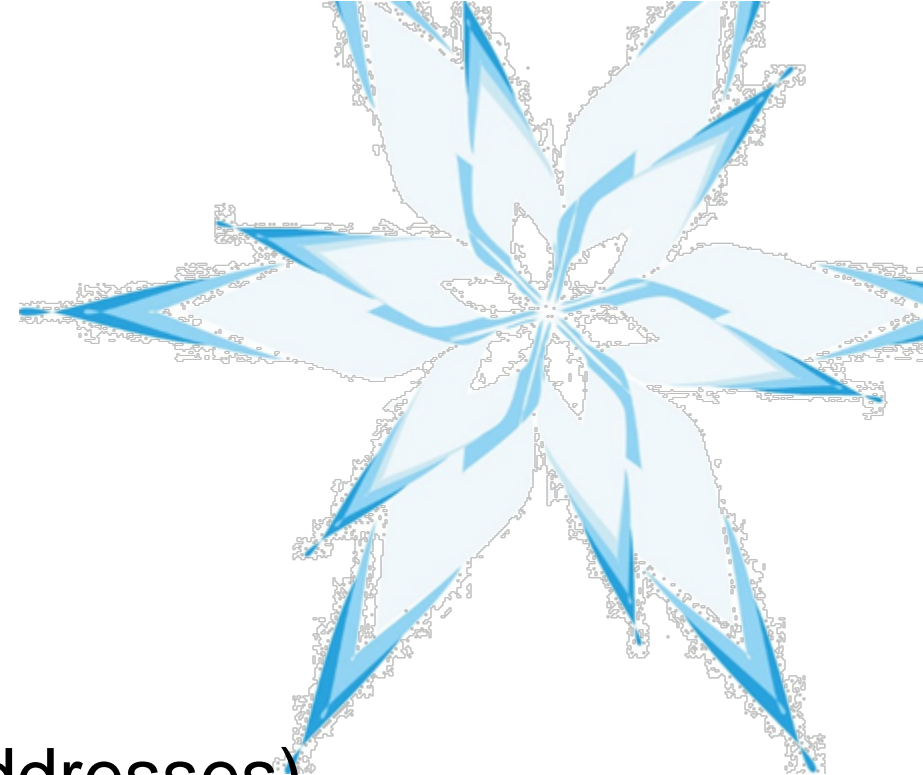
- dc.date.available
- dc.date.issued





# Important fields

- Title
- Author
- Date of publication
- Persistent Identifier (usually URN-based addresses)
- Language
- User rights (preferably CC licences)



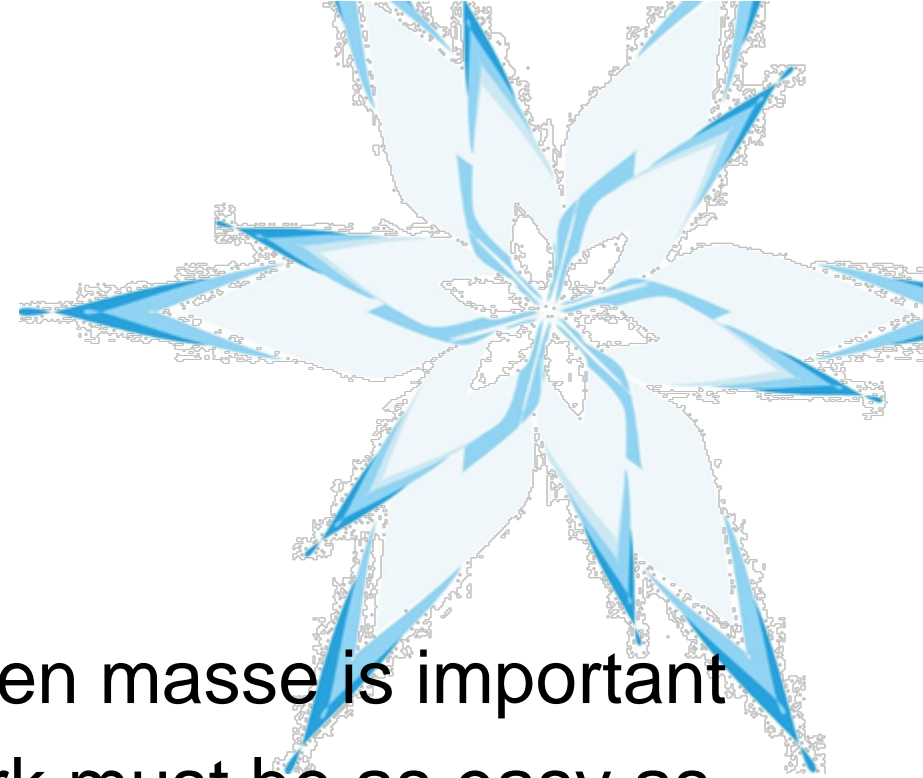
# Future plans and implementation

- Metadata guidelines will be introduced to a wider audience via presentations in seminars and webinars
- Many repositories have made preliminary inspections on how to implement the guidelines
- At the repositories hosted by NatLib alterations to existing records should start in the autumn (after a DSpace version upgrade)
- Guidelines will be kept up to date in the future



# Final thoughts

- Uniform metadata is essential
- Getting the repository community involved en masse is important
- Ways to participate in the development work must be as easy as possible



# We did it, and received



**The Saimaa-Ringed-Seal-of-Approval**



# Thank you!

Metadata guidelines: <https://www.kiwi.fi/x/94R7B>

Tanja Vienonen, [tanja.vienonen@helsinki.fi](mailto:tanja.vienonen@helsinki.fi)

 @tanjamarimint