

# Teel ette tulnud statistikat

Alljärgnevalt mõned statistilised andmed, mis tekkisid andmete töötlemise käigus.

**Viimati uuendatud:** 15. detsember 2025

## Koguarvud

- **Koguarv kirjeid ENTUs:** **41,913** (CSV [entust/eksponaat.csv](#))
- **Faili ridu kokku:** 80,179 rida (1 päis + 41,913 kirjet + 38,265 põimitud reavahetust)
- **Väljade arv kirje kohta:** 44 välja
- **Unikaalseid koode:** 41,887 (41,896 kirjel on kood)
- **Persoone/organisatsioone:** 2,438 isikut + 2 organisatsiooni

Märkus: Põimitud reavahetused

CSV fail sisaldab **38,265 põimitud reavahetust** tsiteeritud väljade sees. See selgitab erinevust 80,179 rea ja 41,913 kirje vahel.

**Top 5 väljad kõige rohkemate reavahetustega:**

1. [photo\\_orig](#): 12,268 reavahetust (pildi URL-id, üks rea kohta)
2. [photo](#): 11,869 reavahetust (pildi URL-id)
3. [repr\\_memento](#): 3,414 reavahetust (represseerimise mälestused)
4. [m2rks6nad](#): 2,280 reavahetust (märkused/kommentaarid)
5. [public\\_legend](#): 1,810 reavahetust (avalikud kirjeldused)

## Kuuluvuste jaotus

Kuuluvus	Arv	Protsent
Dokumendikogu	10247	12.78%
Ajaloolised esemed	7159	8.93%
Kasutuskogu	5241	6.54%
Fotokogu	3990	4.98%
Pisitrükised	2856	3.56%
Maha kantud	2591	3.23%
Arhiivraamatukogu	2223	2.77%
Tekstiilikogu	2219	2.77%
Märgikogu	2169	2.71%
Plakatikogu	918	1.14%

Kuuluvus	Arv	Protsent
Auviste kogu	603	0.75%
Numismaatikakogu	537	0.67%
Kunstikogu	358	0.45%
Digifotokogu	349	0.44%
Militaaria	338	0.42%
Digidokumendikogu	103	0.13%
Ajaloolised esemed	4	0.00%

**Kokku:** 41,912 kirjet kuuluvusega (1 kirje ilma kuuluvuseta)

## Andmete täielikkus

### Põhiväljad

Väli	Täidetud kirjeid	Protsent
<code>code</code>	41,896	99.96%
<code>description</code>	41,747	99.60%
<code>donator</code>	41,623	99.31%
<code>asukoht</code>	40,463	96.54%
<code>year</code>	38,054	90.79%
<code>photo_orig</code>	36,997	88.27%
<code>photo</code>	36,583	87.28%
<code>date</code>	35,986	85.85%
<code>dimensions</code>	34,141	81.46%
<code>koht</code>	33,525	79.99%
<code>public_legend</code>	4,284	10.22%
<code>legend</code>	1,392	3.32%
<code>autor</code>	875	2.09%

### Legendväljad (Issue #14)

#### Legend-väljad lisatud MUIS eksporti (v2.2):

- **Column 90:** `public_legend` (avalik legend) - nähtav avalikkusele
- **Column 91:** `legend` (mitteavaliku legend) - sisemiseks kasutuseks

Väli	Täidetud kirjeid	Protsent	Märkus
<code>public_legend</code>	4,284	10.22%	Avalik informatsioon
<code>legend</code>	1,392	3.32%	Sisemised märkused
Mõlemad	209	0.50%	Kirjed kus mõlemad väljad on
Vähemalt üks	5,467	13.04%	Kirjed vähemalt ühe väljaga

#### Kasutusjuhtumid:

- Kogu hooldaja on jaganud info erinevate väljade vahel
- `description` - peamine kirjeldus
- `public_legend` - avalik legend (näitustel kasutatav tekst)
- `legend` - mitteavalik legend (konserveerimise märkused, päritolu info)

#### Isikud ja organisatsioonid (Phase 1)

Eraldatud andmed (`scripts/extract_person_names.py`, 15. detsember 2025):

Kategooria	Arv	Protsent
Unikaalseid isikuid	2,438	99.92%
Unikaalseid organisatsioone	2	0.08%
<b>Kokku entiteete</b>	<b>2,440</b>	<b>100%</b>
Kirjeid isikute/org-dega	48,883	61.0% (kogust)

#### Väljade jaotus:

Väli	Kirjeldus	Hinnanguline arv
<code>donator</code>	Annetajad	~40,000+
<code>autor</code>	Autorid/loojad	~10,000+
<code>represseeritu_o</code>	Represseeritud isikud (ohvrid)	~2,000+
<code>represseeritu_t</code>	Represseeritud isikud (seotud)	~2,000+

#### Kõige sagedamini esinevad:

1. **Heiki Ahonen** - 13,944 kirjet (17.4% andmebaasist)
2. **Aime Pärnakivi** - 1,740 kirjet (2.2%)
3. **Arvo Pesti** - 1,324 kirjet (1.7%)
4. **Ülle Michelson** - 1,102 kirjet (1.4%)
5. **Eino Sandström** - 774 kirjet (1.0%)

**Top 10 hõlmab:** 21,257 esinemist (43.5% kõigist isikute/org viidetest)

#### Tuvastatud organisatsioonid:

1. **Rannarahva Muuseum** - 32 kirjet (Rannarahva muuseum)
2. **Unitas MTÜ** - 12 kirjet (mittetulundusühing)

**Väljund:** `output/person_registry_request.csv` ja `output/person_registry_request.xlsx`  
(2,441 rida) - valmis MuS-i koordineerimiseks

**Järgmised sammud:** Fail saadetakse MuS-i sidusisikule, kes lisab `muis_participant_id` veeru ja tagastab registreerimise järel.

## Pildid

- **Kirjeid piltidega:** 36,997 (88.27%)
- **Pildi URL-e kokku:** 49,265 (mõnel kirjel mitu pilti)
- **Keskmine pilte kirje kohta:** 1.33 pilti

## Kuupäevad

- **Kirjeid kuupäevaga (`date`):** 35,986 (85.85%)
- **Kirjeid aastaga (`year`):** 38,054 (90.79%)
- **Ainult aastaga (ilma `date`-ta):** 2,068 kirjet

## Tehnilised märkused

### CSV sisend failistruktuur

- **Välju kokku:** 44 välja kirje kohta
- **Kodeering:** UTF-8
- **Failide suurus:** 29.74 MB
- **Põimitud reavahetused:** 38,265 reavahetust 27 erineva välja sees

### Koodide unikaalsus

- **Kirjeid koodiga:** 41,896
- **Unikaalseid koode:** 41,887
- **Duplikaatkoodid:** 9 koodi (mõnel koodil mitu kirjet)

### Duplikaatkoodide nimekiri:

1. `018906/000` - 2 kirjet
2. `017692/047` - 2 kirjet
3. `019623/000` - 2 kirjet
4. `kasutuskogu` - 2 kirjet
5. `021521/001` - 2 kirjet
6. `021521/002` - 2 kirjet
7. `021521/003` - 2 kirjet
8. `021522/001` - 2 kirjet
9. `021522/002` - 2 kirjet

## Duplikaatkoodide detailne analüüs

Allpool on kõigi duplikaatkoodide detailne ülevaade koos ENTU linkidega ja võtmeväljade võrdlusega.

Kood	Kirje 1	Kirje 2	Peamised erinevused
018906/000	<a href="#">5422b150bdf9203d66cc9291</a> <b>Nimi:</b> 018906/000 /kunst/maal <b>Kirjeldus:</b> Maal, maalil kujutatud tehnikadoktor Johannes Hint <b>Kuuluvus:</b> Kunstikogu <b>Annetaja:</b> Veljo Vask	<a href="#">542e678dbdf9203d66cc93a8</a> <b>Nimi:</b> 018906/000 /dokument/foto <b>Kirjeldus:</b> Foto Johannes Kaivust <b>Kuuluvus:</b> Fotokogu <b>Annetaja:</b> Peeter Kaiv	<b>Erinev objekt:</b> Üks on maal (Kunstikogu), teine foto (Fotokogu). Erinevad annetajad. Tõenäoliselt viga koodi määramisel - sama kood kasutatud kahele erinevale objektile.
017692/047	<a href="#">56715a31bdf9203d66cca05f</a> <b>Nimi:</b> 017692/047 /dokument/kiri <b>Kirjeldus:</b> 30.06.1967 Johannes Lindermannilt, vennasele <b>Kuuluvus:</b> Dokumendikogu <b>Annetaja:</b> Laine Reinberg	<a href="#">56715a32bdf9203d66cca078</a> <b>Nimi:</b> 017692/047 /dokument/kiri <b>Kirjeldus:</b> 04.02.1971 Johannes Lindermannilt <b>Kuuluvus:</b> Dokumendikogu <b>Annetaja:</b> Laine Reinberg	<b>Erinevad dokumendid:</b> Kaks erinevat kirja samalt saatjalt, erinevatel kuupäevadel (1967 vs 1971). Sama kood kasutatud seeria kirjadele.
019623/000	<a href="#">56d8094dbdf9203d66cca1c1</a> <b>Nimi:</b> 019623/000 /dokument/order <b>Kirjeldus:</b> Tallinna Linna TSN TK Elamispinna arvestuse ja jaotamise osakonna order 13.06.1973 <b>Kuuluvus:</b> Dokumendikogu <b>Annetaja:</b> Jüri Pert	<a href="#">56d80aabbdf9203d66cca1c2</a> <b>Nimi:</b> 019623/000 /dokument/tunnistus <b>Kirjeldus:</b> Ratsionaliseerimisettepanekute tunnistused, venekeelsed, 6 tk <b>Kuuluvus:</b> Dokumendikogu <b>Annetaja:</b> Jüri Pert	<b>Erinevad dokumendid:</b> Üks on order, teine tunnistused. Erinevad dokumenditüübid, sama annetaja. Tõenäoliselt viga koodi määramisel.
kasutuskogu	<a href="#">58088722bdf9203d66cca4e4</a> <b>Nimi:</b> kasutuskogu /varia/illustreeriv materja <b>Kirjeldus:</b> Seinalehe tarbeks propagandistlikud fotod ja trükitud materjalid Leninist <b>Kuuluvus:</b> Kasutuskogu <b>Annetaja:</b> Heldi Veedler	<a href="#">583fc9a8bdf9203d66cca658</a> <b>Nimi:</b> kasutuskogu /dokument/liikmepilet <b>Kirjeldus:</b> Komsomolipilet, Raissa Ploštšik <b>Kuuluvus:</b> Kasutuskogu <b>Annetaja:</b> Heldi Veedler	<b>Viga:</b> "kasutuskogu" on kasutatud koodina, mitte kuuluvusena. Mõlemad on Kasutuskogus, kuid puudub õige kood. See on selge andmete sisestamise viga.

Kood	Kirje 1	Kirje 2	Peamised erinevused
021521/001	<a href="#">6270a25dbdf9203d66ccb3fe</a> <b>Nimi:</b> 021521/001 /dokument/foto <b>Kirjeldus:</b> Fotokoopia metsavennast Olev Turust koos emaga <b>Kuuluvus:</b> Fotokogu <b>Annetaja:</b> Krete Kruusmaa	<a href="#">63105193bdf9203d66ccb550</a> <b>Nimi:</b> 021521/001 <b>Kirjeldus:</b> /aparaat/föön <b>Kuuluvus:</b> Maha kantud <b>Annetaja:</b> (tühi)	<b>Süsteemne duplikatsioon:</b> Üks on aktiivne foto Fotokogus, teine on maha kantud objekt (föön). Sama kood kasutatud erinevate objektidele - võimalik, et kood taaskasutatud pärast esimese objekti mahakandmist.
021521/002	<a href="#">6270a2b6bdf9203d66ccb3ff</a> <b>Nimi:</b> 021521/002 /dokument/foto <b>Kirjeldus:</b> Fotokoopia metsavennast Olev Turust koos vanematega <b>Kuuluvus:</b> Fotokogu <b>Annetaja:</b> Krete Kruusmaa	<a href="#">63105194bdf9203d66ccb551</a> <b>Nimi:</b> 021521/002 <b>Kirjeldus:</b> /aparaat/föön <b>Kuuluvus:</b> Maha kantud <b>Annetaja:</b> (tühi)	<b>Süsteemne duplikatsioon:</b> Sama muster - aktiivne foto vs maha kantud objekt.
021521/003	<a href="#">6270a2d6bdf9203d66ccb400</a> <b>Nimi:</b> 021521/003 /dokument/foto <b>Kirjeldus:</b> Fotokoopia metsavennast Olev Turust koos vanematega <b>Kuuluvus:</b> Fotokogu <b>Annetaja:</b> Krete Kruusmaa	<a href="#">63105194bdf9203d66ccb552</a> <b>Nimi:</b> 021521/003 <b>Kirjeldus:</b> /aparaat/elektronlamp <b>Kuuluvus:</b> Maha kantud <b>Annetaja:</b> (tühi)	<b>Süsteemne duplikatsioon:</b> Sama muster - aktiivne foto vs maha kantud objekt (elektronlamp).
021522/001	<a href="#">6270a2f7bdf9203d66ccb401</a> <b>Nimi:</b> 021522/001 /dokument/foto <b>Kirjeldus:</b> Foto metsavennast Olev Turust koos vanematega <b>Kuuluvus:</b> Fotokogu <b>Annetaja:</b> Krete Kruusmaa	<a href="#">63105194bdf9203d66ccb553</a> <b>Nimi:</b> 021522/001 <b>Kirjeldus:</b> /aparaat/elektronlamp <b>Kuuluvus:</b> Maha kantud <b>Annetaja:</b> (tühi)	<b>Süsteemne duplikatsioon:</b> Sama muster - aktiivne foto vs maha kantud objekt.

Kood	Kirje 1	Kirje 2	Peamised erinevused
021522/002	<p>6270a321bdf9203d66ccb402</p> <p><b>Nimi:</b> 021522/002 /dokument/foto</p> <p><b>Kirjeldus:</b> Foto metsavennast Olev Turust koos emaga</p> <p><b>Kuuluvus:</b> Fotokogu</p> <p><b>Annetaja:</b> Krete Kruusmaa</p>	<p>63105195bdf9203d66ccb554</p> <p><b>Nimi:</b> 021522/002</p> <p><b>Kirjeldus:</b> /aparaat/elektronlamp</p> <p><b>Kuuluvus:</b> Maha kantud</p> <p><b>Annetaja:</b> (tühi)</p>	<p><b>Süsteemne duplikatsioon:</b></p> <p>Sama muster - aktiivne foto vs maha kantud objekt.</p>

## Duplikaatide põhjused ja soovitused

### Kategooriad:

#### 1. Koodi taaskasutus pärast mahakandmist (021521/XXX, 021522/XXX):

- 6 koodi, kus sama kood on kasutatud nii aktiivsele objektile (Fotokogu) kui ka maha kantud objektile
- Tõenäoliselt on koodid taaskasutatud pärast esimese objekti mahakandmist
- **Soovitus:** Kontrollida, kas koodide taaskasutus on lubatud või peaks olema keelatud

#### 2. Vale koodi määramine (018906/000, 019623/000):

- Sama kood kasutatud erinevate objektidele
- Erinevad kuuluvused või dokumenditüübid
- **Soovitus:** Üks kirjeid vajab uut koodi

#### 3. Seeria kirjete koodid (017692/047):

- Kaks erinevat kirja samalt saatjalt, erinevatel kuupäevadel
- **Soovitus:** Võib olla õige, kui on tegu seeriaga, kuid võiks kasutada järjenumbreid (nt 017692/047-1, 017692/047-2)

#### 4. Andmete sisestamise viga (kasutuskogu):

- "kasutuskogu" on kasutatud koodina, mitte kuuluvusena
- **Soovitus:** Mõlemad kirjed vajavad õigeid koode

**Kokkuvõte:** 9 duplikaatkoodi, millest 6 on süsteemne probleem (koodide taaskasutus), 2 on tõenäoliselt viga, ja 1 võib olla õige seeria kood.