

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

Top 5 väljad kõige rohkemate reavahetustega:

- [photo_orig](#): 12,268 reavahetust (pildi URL-id, üks rea kohta)
- [photo](#): 11,869 reavahetust (pildi URL-id)
- [repr_memento](#): 3,414 reavahetust (represseerimise mälestused)
- [m2rks6nad](#): 2,280 reavahetust (märkused/kommentaarid)
- [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	41,896	99.96%
description	41,747	99.60%
donator	41,623	99.31%
asukoht	40,463	96.54%
year	38,054	90.79%
photo_orig	36,997	88.27%
photo	36,583	87.28%
date	35,986	85.85%
dimensions	34,141	81.46%
koht	33,525	79.99%
public_legend	4,284	10.22%
legend	1,392	3.32%
autor	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
public_legend	4,284	10.22%	Avalik informatsioon
legend	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%
Kokku entiteete	2,440	100%
Kirjeid isikute/org-dega	48,883	61.0% (kogust)

Väljade jaotus:

Väli	Kirjeldus	Hinnanguline arv
donator	Annetajad	~40,000+
autor	Autorid/loojad	~10,000+
represseeritu_o	Represseeritud isikud (ohvrid)	~2,000+
represseeritu_t	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 MuIS-i koordineerimiseks

Järgmised sammud: Fail saadetakse MuIS-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	<p>5422b150bdf9203d66cc9291</p> <p>Nimi: 018906/000 /kunst/maal</p> <p>Kirjeldus: Maal, maalil kujutatud tehnikadoktor Johannes Hint</p> <p>Kuuluvus: Kunstikogu</p> <p>Annetaja: Veljo Vask</p>	<p>542e678dbdf9203d66cc93a8</p> <p>Nimi: 018906/000 /dokument/foto</p> <p>Kirjeldus: Foto Johannes Kaivust</p> <p>Kuuluvus: Fotokogu</p> <p>Annetaja: Peeter Kaiv</p>	<p>Erinev objekt: Üks on maal (Kunstikogu), teine foto (Fotokogu).</p> <p>Erinevad annetajad.</p> <p>Tõenäoliselt viga koodi määramisel - sama kood kasutatud kahele erinevale objektile.</p>
017692/047	<p>56715a31bdf9203d66cca05f</p> <p>Nimi: 017692/047 /dokument/kiri</p> <p>Kirjeldus: 30.06.1967 Johannes Lindermannilt, vennasele</p> <p>Kuuluvus: Dokumendikogu</p> <p>Annetaja: Laine Reinberg</p>	<p>56715a32bdf9203d66cca078</p> <p>Nimi: 017692/047 /dokument/kiri</p> <p>Kirjeldus: 04.02.1971 Johannes Lindermannilt</p> <p>Kuuluvus: Dokumendikogu</p> <p>Annetaja: Laine Reinberg</p>	<p>Erinevad dokumendid: Kaks erinevat kirja samalt saatjalt, erinevatel kuupäevadel (1967 vs 1971). Sama kood kasutatud seeria kirjadele.</p>
019623/000	<p>56d8094dbdf9203d66cca1c1</p> <p>Nimi: 019623/000 /dokument/order</p> <p>Kirjeldus: Tallinna Linna TSN TK Elamispinna arvestuse ja jaotamise osakonna order 13.06.1973</p> <p>Kuuluvus: Dokumendikogu</p> <p>Annetaja: Jüri Pert</p>	<p>56d80aabbd9203d66cca1c2</p> <p>Nimi: 019623/000 /dokument/tunnistus</p> <p>Kirjeldus: Ratsionaliseerimisetepaneküte tunnistused, venekeelsed, 6 tk</p> <p>Kuuluvus: Dokumendikogu</p> <p>Annetaja: Jüri Pert</p>	<p>Erinevad dokumendid: Üks on order, teine tunnistused.</p> <p>Erinevad dokumenditüübidi, sama annetaja.</p> <p>Tõenäoliselt viga koodi määramisel.</p>
kasutuskogu	<p>58088722bdf9203d66cca4e4</p> <p>Nimi: kasutuskogu /varia/illustreeriv materja</p> <p>Kirjeldus: Seinalehe tarbeks propagandistlikud fotod ja trükitud materjalid Leninst</p> <p>Kuuluvus: Kasutuskogu</p> <p>Annetaja: Heldi Veedler</p>	<p>583fc9a8bdf9203d66cca658</p> <p>Nimi: kasutuskogu /dokument/liikmepilet</p> <p>Kirjeldus: Komsomolipilet, Raissa Ploštšik</p> <p>Kuuluvus: Kasutuskogu</p> <p>Annetaja: Heldi Veedler</p>	<p>Viga: "kasutuskogu" on kasutatud koodina, mitte kuuluvusena.</p> <p>Mõlemad on Kasutuskogus, kuid puudub õige kood. See on selge andmete sisestamise viga.</p>

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

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

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 dokumenditüübide
- **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 järgenumbreid (nt 017692/047-1, 017692/047-2)

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

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

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