

Teel ette tulnud statistikat

Alljärgnevalt mõned statistilised andmed, mis tekkisid andmete töötlemise käigus.

Viimati uuendatud: 11. 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)

Märkus: Põimitud reavahetused

CSV fail sisaldab **38,265 põimitud reavahetust** tsiteeritud väljade sees. See selgitab erinevust 80,179 rea ja 41,913 kirje vahel. Python'i [csv.DictReader](#) töötleb need õigesti.

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/kommentaariid)
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%
Auviste kogu	603	0.75%

Kuuluvus	Arv	Protsent
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>autor</code>	875	2.09%

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 failistruktuur

- **Välju kokku:** 44 välja kirje kohta
- **Kodeering:** UTF-8
- **Faili 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 (tõenäoliselt viga)
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

Märkus: Seeria 021521/XXX ja 021522/XXX (6 koodi) näib olevat süsteemne duplikatsioon - võimalik, et sama objekt erinevates registrites.