

# Az OSZK webaratás pilot projektjének gyűjtőköri tervezete

Utolsó módosítás: 2017.06.02.

*A tervezet szövege a pilot időszak vége (2018 decembere) után felülvizsgálásra kerül majd.*

Az Országos Széchényi Könyvtár [gyűjtőköri szabályzatában](#) rögzített feladatai közé tartozik a magyar webes tartalmak gyűjtése, reprezentatív képet nyújtva ily módon a múltban egy adott időpillanatban elérhető, a magyar közönségnek szánt, a kulturális örökség részét képező online tartalomkínálatról, a hungarikumok körébe tartozó és nyilvános szolgáltatási felületekről letölthető elektronikus dokumentumokról.

A dokumentumok gyűjtése háromféle módon történik: válogatva a legfontosabb magyar webhelyekről, kiemelt eseményekhez kötődve a főbb hírforrásokból, illetve általános jelleggel a magyar webtérrel. Szelektíven kerül gyűjtésre a tudományos, kulturális, oktatási, közéleti jellegű tartalmak meghatározott köre. Az általános gyűjtés a .hu domén alatt regisztrált, vagy egyéb doménhez tartozó, de magyar közönséget megcélzó nyilvános webhelyekre terjed ki, az illegálisnak illetve pornográfnek minősülő tartalmak kiszűrésével, valamint a nagyméretű multimédia és egyéb állományok kihagyásával.

A webaratás csupán azon webhelyeket érinti, ahonnan technikailag biztosítható a tartalom automatikus begyűjtése. Az aratási művelet során a könyvtár figyelembe veszi a begyűjtő robot számára az adott webhely tulajdonosa által beállított korlátozásokat (robots.txt).

Az archivált webhelyek esetében a könyvtár elsősorban azok tartalmának hosszú távú megőrzésére törekszik. A begyűjtött tartalom szolgáltatása során teljes mértékben garantálni kell, hogy a jogtulajdonosok szellemi tulajdonjoga ne sérüljön. Nyilvános szolgáltatás csak azokról a webhelyekről végezhető, amelyek esetében a tartalomgazda erre engedélyt adott, a többi szelektíven archivált honlap csak a könyvtár zárt szolgáltatási felületén érhető el, az általános webtér aratás anyaga pedig a törvényi szabályozás megszűletéséig nem hozzáférhető.