

## [Tematická kolekce webových zdrojů COVID-19 jako součást Webarchivu \[1\]](#)

### **Autor:**

- [HAŠKOVCOVÁ, Marie](#) [2], [VOZÁR, Zdenko](#) [3]

### **Číslo:**

- [2020, ročník 29, číslo 1](#) [4]

### **Rubrika:**

- [Trendy v knihovnách](#) [5]

### **Klíčová slova:**

- [archivace webu](#) [6], [doba koronavirová](#) [7]

Letošní rok jsme se v důsledku pandemie viru covid-19 ocitli v nebývalé situaci. Virus zasáhl do našich životů na mnoha úrovních, a proto se [Český webový archiv](#) [8] (Webarchiv), podobně jako řada dalších zahraničních archivů, pokusil postihnout dění v prostředí internetu související s šířením a dopadem viru ve speciální tematické kolekci elektronických zdrojů.

Webarchiv se coby součást Národní knihovny ČR snaží zachytit vzorek proměnlivého dění v prostředí internetu, shromažďovat, uchovat i zpřístupňovat informace, které představují specifickou součást našeho kulturního dědictví. Akviziční politika kombinuje tzv. sklizně celoplošné (sběr celé domény .cz jednou až dvakrát do roka), výběrové (dlouhodobě budované kolekce pravidelně sklízených bohemikálních zdrojů řazených do jednotlivých předmětových kategorií na základě metody Konspektu) a tematické (kolekce sledující určité téma). Podrobně se budování sbírek Webarchivu věnuje dokument [Strategie budování sbírky Webarchivu](#) [9]. Jedna z tematických kolekcí se zaměřila právě na [covid-19](#) [10].

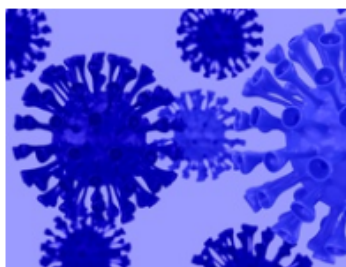
Jaké zdroje kolekce zahrnuje? Snažili jsme se o co nejkomplexnější zachycení přítomnosti a dopadu viru. Zařazeny do ní byly vládní weby, odborné vědecké i popularizační zdroje, ohlasy v médiích, stránky dobrovolnických aktivit a občanských iniciativ, akademických pracovišť i jednotlivců, kteří se věnují vývoji ochranných pomůcek. Součástí kolekce jsou zdroje zkoumající ekonomické, právní a sociální dopady, různé statistiky (publikované státem, firmami nebo jednotlivci) a matematické modely i diskuse k nim (např. na GitHub), ale třeba i umělecké reflexe přítomnosti koronaviru nebo různé dezinformace, které se v souvislosti s covidem-19 objevily. Vzhledem k výjimečnosti situace – byl vyhlášen nouzový stav a celostátní karanténa – i s ohledem na nepředvídatelnost vývoje situace a na měnící se vládní opatření jsme se rozhodli tuto kolekci od poloviny března poprvé sklízet na denní bázi. S květnovým uvolněním vládních nařízení a s ukončením plošné karantény jsme se vrátili k měsíčnímu intervalu sklizení kolekce. Podle toho, jak se bude situace vyvíjet, budeme případně opět upravovat i frekvenci sklizení.

# COVID-19

Klíčová slova sklizně:

[krizový management](#), [infekční nemoci](#), [epidemie](#), [nemoci](#), [krizové situace](#), [globální problémy](#), [lékařská péče](#), [přenos infekčních nemocí](#)

---



Koronavirus COVID-19 hýbe světem. Průběžně vznikající tematická kolekce sleduje jeho přítomnost a dopady z různých úhlů. Snaží se zachytit v čase vládní zdroje, tematické weby - odborné i popularizační, lékařské a zdravotní portály, ohlasy v různých typech médií.

SZU.CZ

<http://szu.cz/tema/krizove-situace/2019-ncov-novy-koronavirus-wu-chan> [\[archivovaná verze\]](#)

WWW.MZCR.CZ

[http://www.mzcr.cz/dokumenty/informace-pro-obcany-v-souvislosti-s-aktualnim-vyskytem-cinskeho-koronaviru-2019\\_18415\\_1.html](http://www.mzcr.cz/dokumenty/informace-pro-obcany-v-souvislosti-s-aktualnim-vyskytem-cinskeho-koronaviru-2019_18415_1.html) [\[archivovaná verze\]](#)

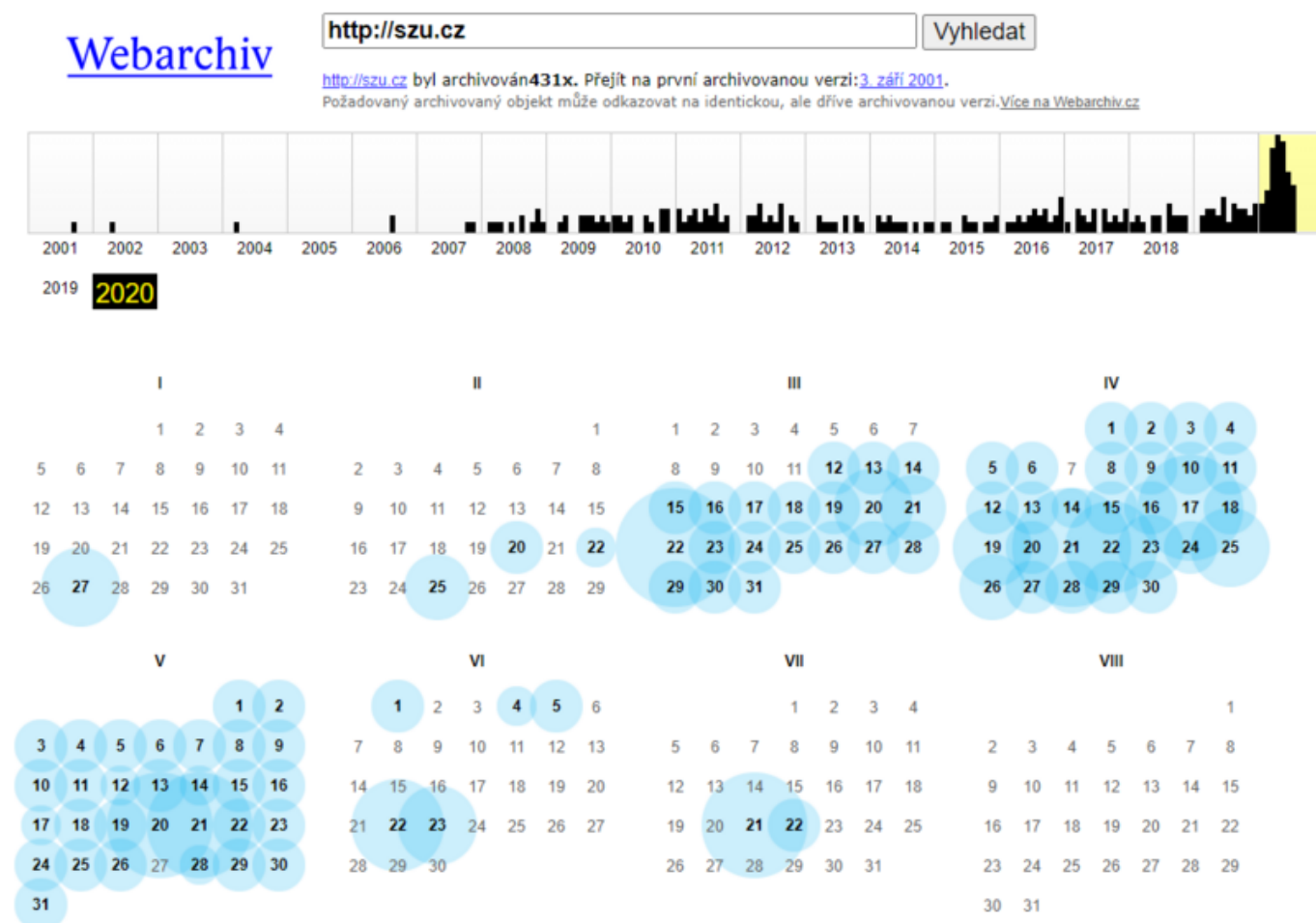
WWW.VLADA.CZ

<https://www.vlada.cz/cz/media-centrum/aktualne/aktualni-informace-ke-koronaviru-2019-ncov-179250/> [\[archivovaná verze\]](#)

[11]

Tematická kolekce COVID-19 na stránkách Webarchivu

Sbírka je výjimečná i v tom, že při akvizici zdrojů byly poprvé použity automatizované postupy. Denní periodicita sklizení (včetně víkendů a svátků) s ideálním začátkem po 19. hodině si vynutila vytvoření rámce pro automatické sklizení tak, aby operátor nemusel spouštět sklizeň ručně. Zároveň automatický start v nastavení zohledňuje změnu rozsahu zdrojů (tj. pokud přibyla od kurátorů nová semínka, tj. URL adresy webových stránek) a upravuje typ názvosloví, např. pro systémové označení jednotlivých balíčků na hierarchickém úložišti. Vzhledem k zaměření sklizně (rychlý snapshot všech zdrojů), maximální délce sklizení (maximálně čtyři hodiny) a poměrně nízkému počtu zdrojů (v rozsahu cca 200 až 550 URL adres) se podařilo pro tuto kolekci získat 2,7 TB dat; v průměru se jednalo přibližně o 30 GB na jednotlivou sklizeň. Ne všechny automatizované sklizně proběhly podle našich představ, ale výsledek je z obsahového i metodologického hlediska dobrým základem do budoucna. Přímou na tuto iniciativu navazuje kontinuální sklizeň [českých médií](#) [12], která internetová periodika a magazíny sklízí dokonce několikrát denně. Zavádění automatizace do postupů českého Webarchivu umožní sbírat zásadní data maximálně pár hodin po zveřejnění, precizovat záběr a kvalitu sběru, zacílit celkové technické parametry konkrétní sklizně a zároveň monitorovat její průběh v reálném čase, což se jeví jako ideální postup pro sledování významných událostí, které se vyvíjejí v čase.



[13]

Záznam sklizní webové stránky <http://szu.cz/> [14] na časové ose – zpřístupňovací aplikace Wayback Machine

Webarchiv se snaží o spolupráci jak s dalšími českými, tak se zahraničními institucemi. Od roku 2007 je členem mezinárodního konsorcia webových archivů IIPC ([International Internet Preservation Consortium](#) [15]). Konsorcium se zaměřuje na sdílení zkušeností, vyvíjení nástrojů pro sklizení webu a formulování tzv. best practices. Jednou z jeho aktivit je i budování společných mezinárodních tematických kolekcí, tzv. [collaborative collections](#) [16]. Iniciovalo i vytvoření velké mezinárodní sbírky věnované covidu-19 – [Novel Coronavirus \(COVID-19\)](#) [17], do které Webarchiv přispěl výběrem českých zdrojů. Zdroje navržené institucemi z celého světa jsou ve spolupráci se službou Archive-It průběžně archivovány od února 2020.

Ve snaze o vytvoření hodnotné sbírky zdrojů jsme oslovili i knihovnickou komunitu a veřejnost s žádostí o zasílání návrhů zdrojů k archivaci prostřednictvím [formuláře](#) [18]. Obdrželi jsme tipy na rozmanité zdroje, včetně knihovnických, zdroje s širším záběrem (jako například portál <https://protiviru.knihovny.cz/> [19]) i zaměřené na ryze regionální témata. Přestože bychom tuto kolekci nejraději uzavřeli, covid-19 s námi zatím stále zůstává. Snažíme se o vytvoření relevantního informačního zdroje pro budoucí bádání, proto i nadále uvítáme tipy zdrojů k archivaci pro rozšíření stávající kolekce.

Obrázky pocházejí z [Webarchivu](#) [20].

**URL zdroje:** <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2020-rocnik-29-cislo-1/tematicka-kolekce-webovych-zdroju-covid-19-jako>

## Odkazy

- [1] <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2020-rocnik-29-cislo-1/tematicka-kolekce-webovych-zdroju-covid-19-jako>
- [2] <https://bulletinskip.skipcr.cz/vsechna-cisla/autori/haskovcova-marie>
- [3] <https://bulletinskip.skipcr.cz/vsechna-cisla/autori/vozar-zdenko>
- [4] <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2020-rocnik-29-cislo-1>
- [5] <https://bulletinskip.skipcr.cz/vsechna-cisla/rubriky/trendy-v-knihovnach>
- [6] <https://bulletinskip.skipcr.cz/vsechna-cisla/klicova-slova/archivace-webu>
- [7] <https://bulletinskip.skipcr.cz/vsechna-cisla/klicova-slova/doba-koronavirova>
- [8] <https://webarchiv.cz/>
- [9] <https://webarchiv.cz/static/www/download/collection-policy.pdf>
- [10] <https://webarchiv.cz/cs/tematicke-kolekce/covid-19>
- [11] <https://bulletinskip.skipcr.cz/sites/default/files/images/728/haskovcova1.png>
- [12] <https://webarchiv.cz/cs/tematicke-kolekce/ceska-media>
- [13] <https://bulletinskip.skipcr.cz/sites/default/files/images/728/haskovcova2.png>
- [14] <http://szu.cz/>
- [15] <https://netpreserve.org/>
- [16] <http://netpreserve.org/projects/collaborative-collections/>
- [17] <https://archive-it.org/collections/13529>
- [18] <https://forms.gle/2tuoALe58Vj6PXXN9>
- [19] <https://protiviru.knihovny.cz/>
- [20] <https://www.webarchiv.cz/>