

10 egyszerű lépés egy replikációs csomag létrehozásához

Miklós Koren¹, Lars Vilhuber², Imola Csóka³, Marie Connolly⁴, Joan Llull⁵

1 - The Review of Economic Studies Data Editor

2 - AEA Data Editor

3 - Research assistant at REStud project

4 - Canadian Journal of Economics Data Editor

5 - Economic Journal and Econometrics Journal Data Editor

2023-04-26

Előadja: **Kiss Gergely Attila**

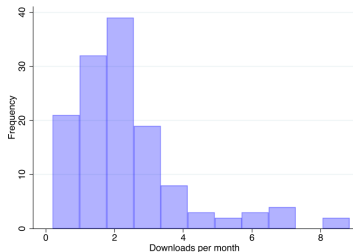
Email: kissg@ceu.edu, gergely.kiss@ksh.hu

DCAS: Data and Code Availability Standard

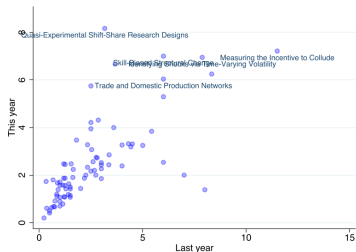
DCAS mint közös irányelv a reprodukálhatóságra

- Fontos előrelépés
- 2022 December 15-e óta publikus
- 5 újság szerkesztői által közösen definiált 16 pontos gyakorlatot ajánl
- Folyóiratok a saját eljárásukat ehhez igazíthatják
- Engedékeny, nem minden részletet kell betartani
- Adaptáló folyóiratok száma is növekedett a kezdeti 5-ről 13-ra.

Vannak-e felhasználók?



(a) Letöltések hisztogramja



(b) Letöltések időbeli korrelációja

Hogyan állítsunk elő jó **replikációs csomagot**?

- Harmonizált irányelvek az folyóiratoknál.
- Saját tapasztalatok a REStudnál (200+ csomag: átlagosan 2,5 forduló)
- Mik a leggyakoribb hibák?

Hogyan állítsunk elő jó **replikációs csomagot**?

- Harmonizált irányelvek az folyóiratoknál.
- Saját tapasztalatok a REStudnál (200+ csomag: átlagosan 2,5 forduló)
- Mik a leggyakoribb hibák?
 - ✗ Hiányos adat dokumentáció.
 - ✗ Szoftververzió, kiegészítő csomag és operációs rendszer kompatibilitások.
 - ✗ Elgépelések, útvonal hibák, véletlenre való kontrollálás a kódban.

Hogyan állítsunk elő jó replikációs csomagot?

- Harmonizált irányelvek az folyóiratoknál.
- Saját tapasztalatok a REStudnál (200+ csomag: átlagosan 2,5 forduló)
- Mik a leggyakoribb hibák?
 - ✗ Hiányos adat dokumentáció.
 - ✗ Szoftververzió, kiegészítő csomag és operációs rendszer kompatibilitások.
 - ✗ Elgépelések, útvonal hibák, véletlenre való kontrollálás a kódban.
- Nem minden tökéletesen replikálható az után sem ha ezek meg vannak: nem megosztható adatok és számítás igényhez köthető problémák.

10 lépésben

- 1 Felhasználó felőli megközelítés

Adat és programkód

- 2 Legyen az adat hozzáférhető
- 3 Adatforrás megjelölés (citálás) és hozzáférhész dokumentálása
- 4 Szoftver és hardver követelmények
- 5 Programkód csatolása

Dokumentáció

- 6 A csomag működése
- 7 Használt táblázatok, ábrák felsorolása
- 8 Minden egyéb használt eszköz csatolása

Végül

- 9 Hozzáférhetőséget biztosító licenz használata
- 10 Csomag újrafuttatása legalább egyszer

1. lépés: Felhasználó felőli megközelítés

Próbáljunk meg empatikusak lenni:

- A szükséges információk minél explicitebb közölése
- A telepítéssel dolgozó megoldások kerülése (operációs rendszerek miatt).
- Replikátoroktól érkező interakció minimalizálása.
- Legalap vetőbb tudást feltételezzünk csak.

2-5. lépés: Adat és programkód

- ② Legyen az adat hozzáférhető: Nyers adat csatolása (ha megengedett).

2-5. lépés: Adat és programkód

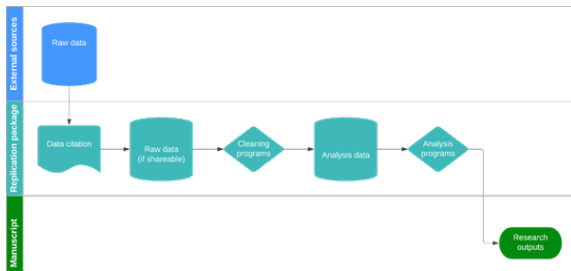
- ② Legyen az adat hozzáférhető: Nyers adat csatolása (ha megengedett).
- ③ Adatforrás megjelölés és hozzáférhész dokumentálása: Adat hozzáférhetőségi és származási nyilatkozat: hogyan, hol és milyen feltételekkel lehet az adathoz hozzáférni.

2-5. lépés: Adat és programkód

- ② Legyen az adat hozzáférhető: Nyers adat csatolása (ha megengedett).
- ③ Adatforrás megjelölés és hozzáférhész dokumentálása: Adat hozzáférhetőségi és származási nyilatkozat: hogyan, hol és milyen feltételekkel lehet az adathoz hozzáférni.
- ④ Szoftver és hardver követelmények : Szoftver verziók, csomagok, operációs rendszer, memória követelmény és futási idő.

2-5. lépés: Adat és programkód

- ② Legyen az adat hozzáférhető: Nyers adat csatolása (ha megengedett).
- ③ Adatforrás megjelölés és hozzáférhős dokumentálása: Adat hozzáférhetőségi és származási nyilatkozat: hogyan, hol és milyen feltételekkel lehet az adathoz hozzáférni.
- ④ Szoftver és hardver követelmények : Szoftver verziók, csomagok, operációs rendszer, memória követelmény és futási idő.
- ⑤ Programkód csatolása: tisztítás, elemzés, ábrák és táblázatok előállítása (Huntington-Klein et al., 2021).



6-8. lépés: Dokumentáció

- ⑥ A csomag működése: README: DAS, követelmények, könyvtár szerkezet, hogyan futtassák, output leírása (Chang and Li, 2017; Krafczyk et al., 2021). Egy fő/rendszerző szkript ajánlott, főleg nagyobb csomagoknál, bonyolultabb eljárásoknál.

6-8. lépés: Dokumentáció

- ⑥ A csomag működése: README: DAS, követelmények, könyvtár szerkezet, hogyan futtassák, output leírása (Chang and Li, 2017; Krafczyk et al., 2021). Egy fő/rendszerző szkript ajánlott, főleg nagyobb csomagoknál, bonyolultabb eljárásoknál.
- ⑦ Táblázatok és ábrák felsorolása: Végtermékek mentése és felsorolása. A programok állítsanak elő mindent!

6-8. lépés: Dokumentáció

- ⑥ A csomag működése: README: DAS, követelmények, könyvtár szerkezet, hogyan futtassák, output leírása (Chang and Li, 2017; Krafczyk et al., 2021). Egy fő/rendszerző szkript ajánlott, főleg nagyobb csomagoknál, bonyolultabb eljárásoknál.
- ⑦ Táblázatok és ábrák felsorolása: Végtermékek mentése és felsorolása. A programok állítsanak elő mindent!
- ⑧ Egyéb használt eszközök (Főleg survey és kísérlet esetén): kérdőív, kitöltési útmutatók, minta kiválasztási részletek.

9-10. lépés: Végül

- 9 Hozzáférhető license: copyright megszüntetése, publikussá válás.

9-10. lépés: Végül

- 9 Hozzáférhető license: copyright megszüntetése, publikussá válás.
- 10 Újra futtatás: teljesen új környezetben. Biztosítja, hogy (a) a csomag minimális erőfeszítéssel reprodukálja az eredményeket és (b) az eredmények megegyeznek.

Hogyan struktúráljunk egy empirikus projektet?

Egy optimális megoldás...

- Minden program & végtermék a Haanwinckel and Soares (2021) cikkhez: [🔗](#)
- A csomag egy Zenodo repozitóriumban [itt](#).

Hasznos anyagok és irányelvek:

- Hatékony gyakorlati tanácsok [itt](#).
- Közös irányelvek a Social Science Data Editors közösségtől [itt](#) újabb gyakorlatok, sablonok, GYIK és folyóirat specifikus követelmények.

A korai erőfeszítések kifizetődnek: áttekinthető projekt könyvtár, kollaboráció elősegítése, másolási hibák kerülése, mások könnyebben tudják újrahasznosítani, gyorsabb felülbírálat.

Konklúziók

- Reprodukálhatóság magával hozza az **áttekinthetőséget**.
- Ajánlott szerkezet:
 - **Nyers adat:** hozzáférhető és citált.
 - **Programok:** transzformációk és eredmények.
 - **Végtermék:** legyen elérhető.
 - **Dokumentáció:** követelmények és lépsek a README fájlban.

Köszönöm a figyelmet!

- Chang, A. C. and Li, P. (2017). A preanalysis plan to replicate sixty economics research papers that worked half of the time. *American Economic Review*, 107(5):60 – 64.
- Haanwinckel, D. and Soares, R. R. (2021). Workforce Composition, Productivity, and Labour Regulations in a Compensating Differentials Theory of Informality. *The Review of Economic Studies*, 88(6):2970–3010.
- Huntington-Klein, N., Arenas, A., Beam, E., Bertoni, M., Bloem, J. R., Burli, P., Chen, N., Grieco, P., Ekpe, G., Pugatch, T., Saavedra, M., and Stopnitzky, Y. (2021). The influence of hidden researcher decisions in applied microeconomics. *Economic Inquiry*, 59(3):944 – 960.
- Kraczyk, M. S., Shi, A., Bhaskar, A., Marinov, D., and Stodden, V. (2021). Learning from reproducing computational results: introducing three principles and the reproduction package . *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 379(2197):20200069.