

IDEGTUDOMÁNYI KUTATÁSI ADATOK ARCHIVÁLÁSA A CONCORDA MAGYAR ADATREPOZITÓRIUMBAN

Fiáth Richárd

ELKH TTK Kognitív Idegtudományi és Pszichológiai Intézet
ORCID: [0000-0001-8732-2691](https://orcid.org/0000-0001-8732-2691)

1. Idegtudományi Kutatások

Szinte triviális kijelentés, hogy az idegtudományi kutatások során keletkező adatok és eredmények rendkívüli értékkel bírnak, ennél fogva ezek hosszú távú, biztonságos megőrzése, illetve nyílt hozzáférésű megosztása a tudományos közösség egyik alapvető fontosságú feladata kell hogy legyen. A központi idegrendszerünk összetett felépítése és működése, illetve az agykutatási adatok létrehozásához használt sokféle különböző kutatási módszer és technológia ezeket az adatokat lényegében egyedivé, pótolhatatlanná teszi.

Mivel módszertani lehetőségek széles tárháza áll rendelkezésünkre, az agykutatáshoz kapcsolódó tudományterületeken számos különféle adattípussal találkozhatunk. A főbb típusok közé tartoznak például a különböző mikroszkópokkal készített anatómiai képek és videók, vagy az agyi elektromos tevékenység vizsgálatára kifejlesztett elektrofiziológiai módszerekkel rögzített adatok, mint például az elektroencefalogram (EEG). A leggyakoribb kutatási adatfajták közé sorolhatók a különféle képalkotó eljárásokkal – például mágneses rezonancia képalkotással (MRI és funkcionális MRI) vagy pozitronemissziós tomográfiával (PET) – készített többdimenziós képek, a kísérletek során keletkezett hang- és videóanyagok (pl. rágszálakon végzett viselkedéses vizsgálatokban), a kognitív tesztek és kérdőívek, vagy az összetettebb idegrendszeri folyamatok szimulálására és modellezésére létrehozott számítási modellek is. De ide tartoznak a nyers adatok feldolgozott és elemzett változatai, vagyis a származtatott adatok, továbbá az adatfeldolgozáshoz és elemzéshez készített és használt szoftverek és forráskódok, illetve az adatokat leíró metaadatok is.

Az agykutatásban használt modern, nagy térbeli és időbeli felbontással rendelkező módszereket alkalmazva manapság már nem ritka, hogy egy kísérlet során több száz gigabájt nyers adat keletkezik. Sőt, egy adott kutatási kérdés vizsgálatára elvégzett hosszabb kísérletsorozat, vagy egy több kutatóintézet együttműködésében megvalósított nagyobb volumenű kutatási projekt esetén akár a petabájtos nagyságrendet is elérheti a projekt életútja során létrejött nyers és származtatott kutatási adat összes mennyisége. Bár az utóbbi példák még viszonylag szélsőséges eseteknek számítanak, és egy átlagos, egy kutatócsoport által megvalósított kutatási projekt esetén a kutatási adatok tárolásához szükséges tárterület általában még nem igényel többet néhány terabájt nál, korunk gyors technológiai fejlődésének eredményeként a létrehozott kutatási adatok mérete is rohamosan növekszik. Ezt a nagy adattömeget egy kutatócsoport – sőt néha egy kutatóintézet is – csak ritkán tudja hosszú távon megőrizni és publikussá tenni. A kutatási pályázatokat kiíró és azokat finanszírozó hazai és nemzetközi támogató szervezetek azonban már gyakran megkövetelik, hogy egy nyertes kutatási projekt során keletkező adatok szabadon hozzáférhetőek legyenek mindenki számára. A kutatási adatok egyszerű és hatékony archiválására és megosztására nyújthatnak ideális megoldást a tudományos adatrepozitóriumok.

Jelen tanulmány célja vázlatos képet adni az idegtudományi kutatási adatok archiválásának folyamatáról és a szerző ezen folyamathoz kapcsolódó tapasztalatairól. Ennek érdekében először a kutatócsoportunkban (Integratív Idegtudományi Kutatócsoport, Kognitív Idegtudományi és Pszichológiai Intézet, Eötvös Loránd Kutatási Hálózat Természettudományi Kutatóközpont)¹ alkalmazott invazívelektrofiziológiai módszerekkel kinyert adatokat (agyí elektromos tevékenység) és ezek főbb tulajdonságait szeretném dióhéjban bemutatni. Ezt követően néhány korábbi kutatási eredményünk alapjait adó elektrofiziológiai mérést, illetve ezen adatok archiválásának és megosztásának folyamatát fogom ismertetni, röviden kitérve a tudományterületen elterjedt főbb szabványos adatformátumokra és az adatok tárolására használt jelentősebb nemzetközi adatrepozitóriumokra.

1 <http://www.ulbertlab.com/>

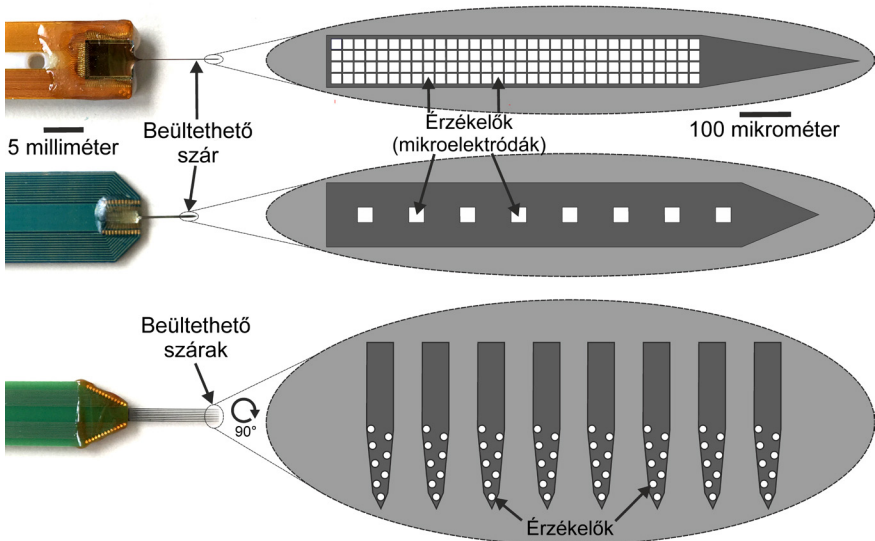
A bemutatott kutatási adatok elhelyezése a közelmúltban kifejlesztett CONCORDA (Concentrated Cooperation on Research Data)² magyar tudományos adatrepozitóriumba történt, egy pilot projekt keretein belül, melyet a Hungarian Research Data Alliance, a Magyar Tudományos Akadémia Könyvtár és Információs Központ, valamint az Eötvös Loránd Kutatási Hálózat (ELKH) Titkársága támogatott.

2. Agyszövetbe ültethető implantátumokkal kinyert kutatási adatok és főbb jellemzőik

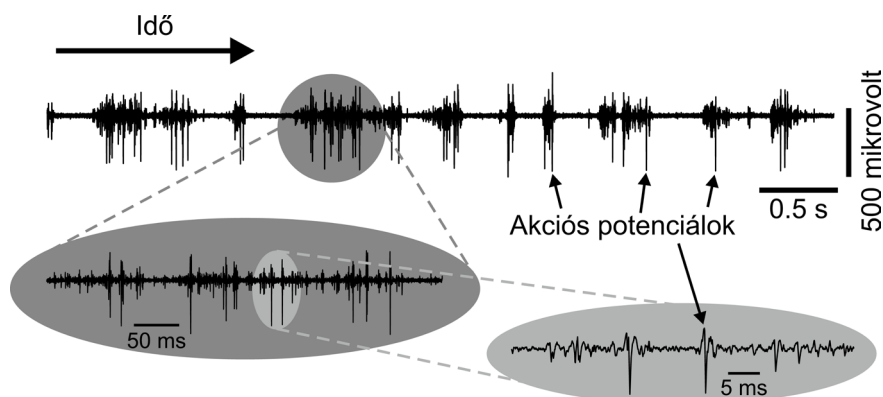
A fő kutatási területem az agyszövetbe ültethető implantátumok, illetve az ilyen típusú eszközökkel rögzíthető agyi elektromos tevékenység vizsgálata, így a továbbiakban ezekre a típusú kutatási adatokra fogok fókuszálni, de néhány esetben kitérek másfajta, idegtudományhoz köthető kutatási adatfajtákra is. A beültethető agyi eszközök alkalmazásának egyik fő célja az eszközön található mikroelektródák (az egyszerűség kedvéért ezeket a továbbiakban érzékelőknek fogom hívni) közelében elhelyezkedő idegsejtek (neuronok) elektromos impulzusainak (szakszóval akciós potenciáljainak) észlelése és rögzítése. Egy ilyen, extracellulárisan (vagyis az idegsejtek közötti térből) mért elektromos impulzus nagyon rövid ideig, nagyjából a másodperc ezredrészéig tart, és ereje (az impulzus alatt mért feszültségváltozás) hozzávetőleg tízezred része egy 1.5 voltos ceruzaelemének. A főként rágcsáló modellben alkalmazott kis méretű agyi implantátumoknak az agyszövetbe ültetett része lényegében egy hajszálvékony (kb. 50 mikrométer vastag), néhány milliméter hosszú tűből (szárból) áll, valamint a tűn található miniatűr (általában kör vagy négyzet alakú) érzékelőkből (1. ábra). Ezekkel az apró érzékelőkkel folyamatosan mérni tudjuk a feszültség változását az agyszövetben, mely feszültséget főként az érzékelők közelében elhelyezkedő idegsejtek tevékenysége befolyásol. Egy adott idegsejt elektromos impulzusa lényegében egy kis tüskéként jelenik meg az extracellulárisan rögzített felvételeken (az angol nyelvű szakirodalomban ezért főként a „spike” kifejezést használják az akciós potenciál szinonimájaként; 2. ábra). Mivel egy-egy érzékelő közelében egyszerre

2 <https://science-data.hu/>

általában sok neuron aktív (úgy is mondhatjuk, hogy „tüzel”), az elektrofiziológias felvételeken számos (sokszor egymással átfedő) túske látható (2. ábra). Egy implantátum általában több érzékelőt is tartalmaz (különböféle geometriai elrendezésben és különböző távolságokra egymástól; lásd az 1. ábrát), a száron elhelyezkedő érzékelők tipikus száma 16 és 32 közé tehető, de a tudományterületen az elmúlt években történt technológiai fejlődésnek köszönhetően ez a szám mára már az ezret is átlépte (Raducanu és mtsai, 2017). Sőt, az egyik legmodernebb, nagy elektróda sűrűségű implantátum négy beültethető szárán összeségében már több mint ötezer ilyen miniatűr érzékelő található (Steinmetz és mtsa, 2021).



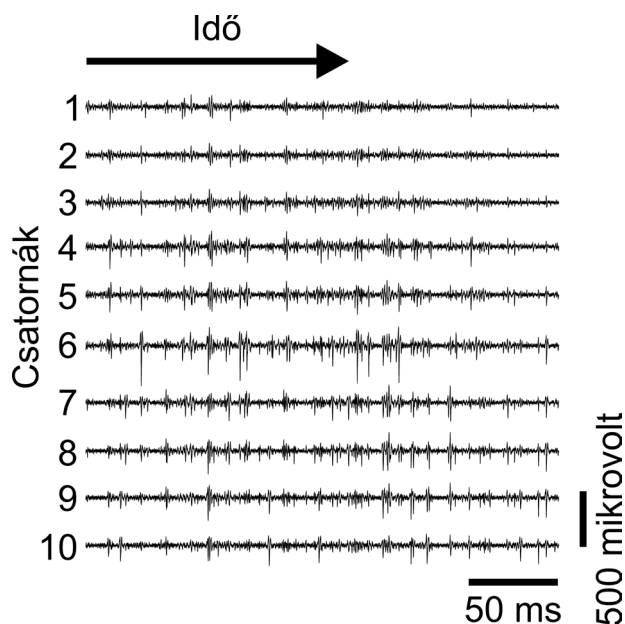
1. ábra: Három különböző, agyszövetbe ültethető implantátum alsó része (balra), valamint az implantátumok szárán található érzékelők (fehér négyzetek és körök) különféle elrendezésben (jobbra). A legelső implantátum nyolc, egymás mellett elhelyezkedő szárral rendelkezik.



2. ábra: Egy érzékélővel mért, több idegsejttől eredő elektromos impulzusok (akciós potenciálok) öt másodperc hosszú mérésben (fent), és ennek egy rövid szakasza kinagyítva, több időskálán (lent). Az elektromos impulzusok kisebb-nagyobb tüskék formájában jelentkeznek a felvételeken (s – másodperc, ms – millimásodperc).

Egy érzékélő csupán a közvetlen környezetében (maximum nagyjából 50–100 mikrométer távolságra) található neuronok akciós potenciálját képes észlelni, így az implantátum szárán található különböző érzékélőkkel, amennyiben azok egymástól viszonylag nagy (>100 mikrométer) távolságra helyezkednek el, különböző idegsejtcsoportok (neuronpopulációk) elektromos tevékenységét fogjuk tudni monitorozni. Ez összességében egyszerre akár több száz, egyedileg azonosítható idegsejt aktivitásának mérését jelenti. Az extracelluláris elektrofiziológiára jellemző tipikus mérési adat tehát a következőképpen áll össze a fentiek alapján: az adatunk egy több csatornából álló mérés, ahol minden csatornán egy adott érzékélőtől eredő idősor látható, vagyis a feszültség időbeli ingadozása, változása az agy egy kicsiny, lokális területén (3. ábra). A mérés során a folytonos (analóg) feszültségjeleket speciális elektronikai áramkörökkel előbb felerősítjük, majd digitalizáljuk. Utóbbi esetén, hogy a nagyon rövid időtartamú elektromos impulzusokat megbízhatóan és jó időbeli felbontással tudjuk rögzíteni, egy adott csatornán másodpercenként legalább húszezerszer mintát veszünk a folytonos jelből, vagyis egy adott érzékélő jeléből ennyi adatpontunk (mért feszültségértékünk) keletkezik egy másodperc alatt. Ez a számérték (húszezer) az úgynevezett mintavételezési frekvencia,

mértékegysége a hertz (Hz). Kissé leegyszerűsítve a folyamatot azt mondhatjuk, hogy a digitális nyers mérési adatokat tulajdonképpen egy táblázatos formában mentjük le a számítógép tárolójára: a táblázat sorai tartalmazzák az egyes csatornákon (érzékelőkön) mért feszültségértékeket, oly módon, hogy a táblázat első oszlopa tartalmazza az összes csatornán mért legelső adatpontot (feszültségértéket), a második oszlop az időben következő (második) adatpontot, és így tovább. A humán EEG mérésekből eredő adatokat is hasonló módon tárolják: általában 32 vagy 64 csatornán (a fejbőrre helyezett elektródák számának függvényében), viszont kisebb mintavételezési frekvenciával (pl. 1000 Hz), mivel az EEG-vel egy másik fajta jeltípust (sok idegsejt közös aktivitásából eredő agyi ritmusokat, lásd lentebb) rögzítünk. Ennek röviden az az oka, hogy mivel az EEG elektródák az idegsejtek méretéhez képest túl nagyok és messze is találhatóak azoktól, az EEG az idegsejtek által kibocsátott elektromos impulzusok közvetlen rögzítésére nem alkalmas.

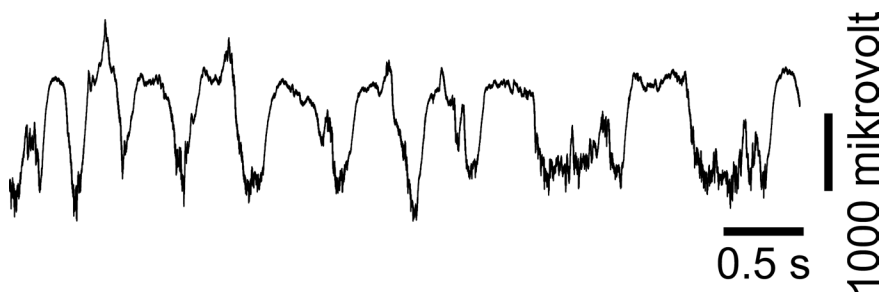


3. ábra: Az ábrán egy 128 érzékelőt tartalmazó agyi implantátummal rögzített agykérgi jel nagyjából negyed másodpercnyi szakasza látható, tíz különböző csatornán (10 érzékelő jele). A felvétel mindegyik csatornáján megfigyelhetők az idegsejtek által kibocsátott, kisebb vagy nagyobb tüskéként jelentkező elektromos impulzusok (ms – millimásodperc).

A sokcsatornás elektrofiziológia területén a nyers kutatási adatok mérete lényegében tehát az alábbiaktól függ: a használt csatornák (érzékelők) száma (minél több, annál nagyobb méretű az adat), a mintavételezési frekvencia (minél többször veszünk mintát egy másodperc alatt, annál nagyobb méretű adatfájl keletkezik), valamint a felvétel hossza (minél hosszabb a mérés, annál több tárterületet foglal). Egy 100 csatornán rögzített, 20 000 Hz-cel mintavételezett, egy óra (3600 mp) hosszúságú mérés például 14.4 GB nagyságú fájl eredményez, amennyiben minden feszültségértéket két bájt nagyságú, egész számokat ábrázoló adattípusként tárolunk el. A tudományterületen általánosak az ennél hosszabb felvételek is, illetve gyakori a magasabb csatornaszámmal rendelkező eszközök használata. Hosszú távra beültetett implantátumokkal például szabadon mozgó rágcsálókból 3–4 óra hosszúságú mérések is megszokottak, akár több mint 300 csatornán, több száz gigabájt nagyságú fájlkat eredményezve. A jelenleg is gőzerővel zajló technológiai fejlesztések eredményeként a következő generációs agyi implantátumokon található érzékelők száma (ezáltal pedig a csatornák száma is) nagy valószínűséggel tovább fog emelkedni, így ennek folyamányaként a kutatási adatok, adatcsomagok méretének jelentős növekedése is várható, hacsak nem kezdenek el időközben olyan tömörítési eljárásokat alkalmazni a területen, melyekkel érdemben csökkenthető a nyers adatok mérete.

Egy adott idegsejt által generált akciós potenciál voltaképpen válasz a hozzá kapcsolódó idegsejtektől rá érkező elektromos impulzusokra. Azonban érdemes megemlíteni egy másik, szintén nagyon fontos jel-típust is, melyet az agyból elektrofiziológiai módszerekkel rögzíteni tudunk: ez pedig az agyi ritmusok (más néven agyhullámok) jelensége. Ahogy fentebb is említettem, a hajas fejbőrre helyezett EEG elektródákkal az egyes neuronok elektromos impulzusait nem tudjuk észlelni, mivel ezek nagyon kis feszültségű és térben gyorsan elhaló események. Az extracelluláris akciós potenciálok csupán néhány tíz mikrométeres távolságra detektálhatóak az érzékelőktől, így tehát az érzékelőnek nagyon közel kell elhelyezkednie egy adott neuronhoz, ha rögzíteni szeretnénk a kibocsátott jeleit. Továbbá az érzékelőknek is nagyon

apróknak kell lenniük, ez nagyjából egy átlagos neuron sejttestének megfelelő méretet (10–20 mikrométer) jelent. Agyunkban egyes idegsejtcsoportok gyakran közel egyszerre, egy időben (szinkron) aktívak, így ezen neuronpopulációk aktivitása (elektromos tere) összeadódik, ami jól elkülöníthető, ritmikusan változó (oszilláló) jelet eredményez a felvételeken (4. ábra). Ezeket az agyi ritmusokat mind a nagyobb méretű EEG elektródákkal (nagy számú, több millió idegsejt összetevékenysége), mind pedig a kisebb, agyszövetbe ültethető implantátumokkal (kisebb neuronpopulációk lokálisabb összetevékenysége) rögzíteni tudjuk. Az egyik, talán legismertebb ilyen agyi ritmus az alfa ritmus, melynek ciklusai (másodpercenként 8–12) jól megfigyelhetők az EEG felvételeken csukott szemű, nyugalmi állapotban lévő embereknél. A beültethető implantátumok érzékelőivel észlelt elektromos impulzusok mellett a kutatók általában az agyi ritmusokat is rögzítik, mivel például érdekes felfedezéseket lehet tenni az agyi ritmusok és a neuronok által generált elektromos impulzusok közötti időbeli összefüggéseket tanulmányozva.



4. ábra: Altatott patkányok agykérgéből rögzített ritmikus agyi tevékenység, az úgynevezett agykérgi lassú ritmus, melynek átlagosan egy másodpercenként jelenik meg egy-egy újabb ciklusa. s – másodperc

A kutatók a nyers mérési adatokat a kísérletek, mérések után a tudományterületre jellemző módszerekkel és algoritmusokkal feldolgozzák, majd elemzik. Ezek főként időtartományban történő elemzéseket jelentenek, mint például annak vizsgálata, hogy milyen időbeli kapcsolat (együttjárás vagy korreláció) figyelhető meg egy idegsejt tüzelési időpontjai

(mikor adott ki elektromos impulzusokat a neuron) és a jutalomadás (pl. cukrozott vizet kap az egér) időpontjai között. Egy másik jelentős vizsgálati módszertan a frekvenciatartományban történő spektrális elemzések, melyet többek között az agyi ritmusok vizsgálatára alkalmaznak, például annak tanulmányozására, hogy milyen agyhullámok alakulnak ki az agy egyes területein az alvás különböző fázisai alatt. A mesterséges intelligencia területén használt módszereket, mint például a mélytanuló algoritmusokat, is egyre gyakrabban alkalmazzák az idegtudományi kutatási adatok feldolgozásához és analizálásához. Jelen tanulmány szűkös keretei között csupán az egyik alapvető, de fontos feldolgozási módszert, a sejt-válogatást (spike sorting) említem meg, mely gyakorlatilag az egyik kezdő lépése az extracellulárisan rögzített, nagyszámú neuron akciós potenciálját tartalmazó nyers adatok feldolgozásának. A sejtvalogatás módszerével tulajdonképpen a sokcsatornás mérésekben található tüskéket rendeljük hozzá egy-egy idegsejthez, vagyis megpróbáljuk meghatározni, hogy melyik elektromos impulzus melyik neurontól eredhetett, mivel erről nincs tudomásunk a mérések során (nem ismerjük az idegsejtek érzékelőkhöz viszonyított helyzetét). Vagy egy másik nézőpontból megközelítve a módszert: annak azonosítása a cél, hogy mely impulzusok származhattak ugyanattól az idegsejttől. Ez az azonosítás általában a különböző idegsejtek elektromos impulzusai közötti kis (időbeli vagy nagyságbeli) különbségek alapján történik: ugyanazon neuron által kibocsátott elektromos impulzusok például nagyjából ugyanolyan hullámformával fognak megjelenni a felvételeken (amennyiben nem mozdítjuk el az érzékelőnket), míg egy szomszédos idegsejt akciós potenciáljai már némileg eltérő hullámformával kerülnek rögzítésre. Ennek a különbségnek a fő oka az, hogy az észlelt akciós potenciál hullámformájának nagysága és alakja nagymértékben függ az idegsejt érzékelőtől való távolságától, illetve az idegsejt típusától is. A sejtvalogatás eljárásának végeredménye a kiválogatott idegsejtek listája lesz (ez néhány sejtől néhány száz neuronig terjedhet felvételenként), valamint ezen idegsejttől eredő elektromos impulzusok megjelenésének időpontjai az adott felvételen. Az adatok további elemzése során gyakran ezeknek az egyedülálló neuronoknak az időbeli tevékenységét viszonyítjuk egymáshoz, vagy a kísérlet egyéb változóihoz, mint például a fentebb említett jutalomadáshoz.

A származtatott kutatási adatok sok esetben kisebb méretűek, mint az eredeti nyers mérési adatok. Erre jó példa a sejtválogatás eredményeként kapott adatok: tegyük fel, hogy a néhány bekezdéssel korábban példaként felhozott egy óra hosszúságú felvételünkben azonosítottunk száz egyedi neuront, és ezek mindegyike átlagosan 5000 elektromos impulzust bocsátott ki az egy órás időintervallum alatt. Mivel mindegyik impulzus egyetlen számértékkel reprezentálható (az akciós potenciál kibocsátásának időpontja a felvétel kezdetéhez képest), a fenti adathalmaz összesen 500 000 adatponttal lesz leírható (ha feltesszük, hogy a mérésünk többi része zaj, és nem tartalmaz számunkra további hasznos információt). Ez így a nyers adathoz képest egy majdnem 15 000-szeres csökkenés az adatpontok számában, vagyis az adat méretében! Ez a méretcsökkenés előnyös lehet, ha például az elérhető szabad tárterület korlátjai miatt a nyers kutatási adatok hosszú távú archiválására, eltárolására nincs lehetőségünk. Ebben az esetben, bár némi információvesztés mellett, de a származtatott adatok eltárolása is egy jó és hasznos alternatíva lehet.

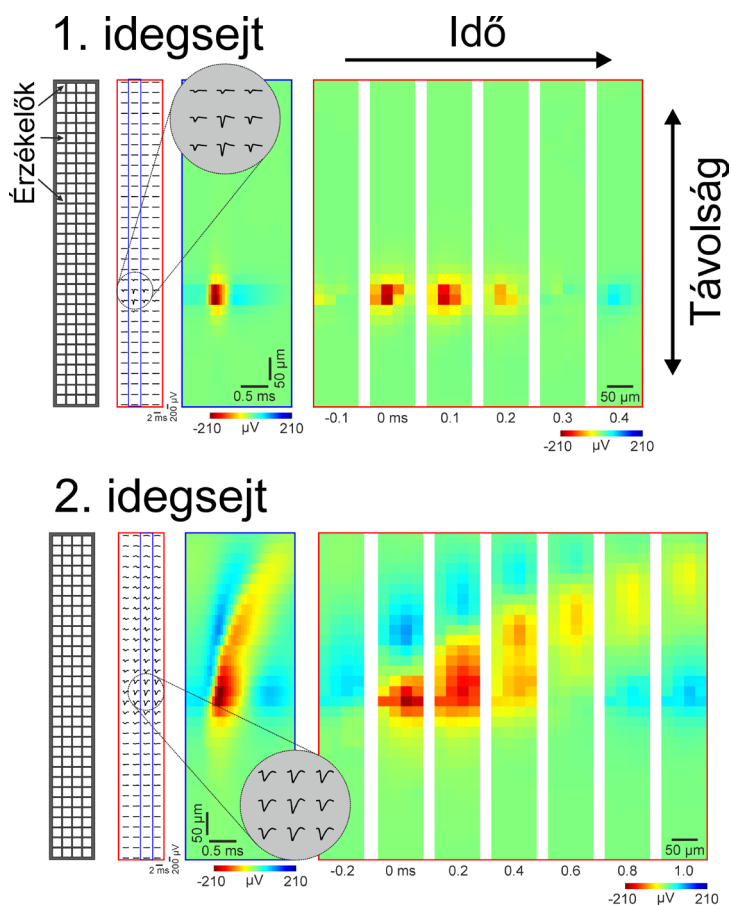
Érdeemes még megemlíteni az adatok egy különleges típusát, a metaadatot. A metaadat lényegében a kutatási adatot leíró információ, megadása nélkül az adatok nehezen értelmezhetők, így sokszor használhatatlannak lennének. Esetünkben a metaadatok közé sorolhatók például az elektrofiziológiai mérések paraméterei: hány csatornán, milyen mintavételezési frekvencia mellett, milyen implantátummal történt a mérés. De metaadat a vizsgált agyterület neve, hogy milyen agyi (sztereotaxiás) koordinátákra és milyen agyi mélységbe ültettük be az implantátumot, vagy, amennyiben valamilyen kísérleti protokollt is alkalmaztunk a kísérletek során, akkor ennek a paraméterei is. Utóbbira egy példa különféle vizuális ingerek alkalmazása a látórendszerhez kapcsolható agyi területek vizsgálatára. Ezen ingereket leíró jellemzők, paraméterek is a metaadatokhoz sorolhatók tehát. Véleményem szerint alapvető fontosságú, hogy lehetőségeinkhez mérten minél több releváns metaadatot mellékeljünk kutatási adataink mellé, hogy szükség esetén akár kollégáink, akár más kutatócsoportok (vagy akár a kísérleteket követően mi magunk is) értelmezni, használni tudják az eltárolt és megosztott

adattípusokat – akár a kutatás eredményeinek reprodukálása céljából, akár új tudományos kérdések vizsgálatára.

3. Az archivált kutatási adatok és az ezekhez kapcsolódó tanulmányok rövid bemutatása

Az Integratív Idegtudományi Kutatócsoportban alkalmazott egyik alapvető kísérleti módszerünk a beültethető agyi implantátumok használata. A kutatócsoport az elmúlt évek során több ilyen típusú, szilícium-alapú implantátum fejlesztésében és tesztelésében is részt vett, illetve rendszeresen használja is ezeket az eszközöket különféle agyi ritmusok (pl. a mélyalvás, altatás során megfigyelhető lassú agykérgi ritmus; lásd a 4. ábrát) vizsgálatára rágcsáló modellen. Jelen tanulmányban az ilyen típusú eszközökkel rögzített sokcsatornás elektrofiziológiai mérések archiválásával és megosztásával kapcsolatos tanulságokat, valamint a szerzett tapasztalataimat szeretném bemutatni. Ezek a kísérleti adatok szélessávú mérések – tehát egyaránt tartalmazzák az agyi ritmusokat és az egyes neuronok elektromos impulzusait is –, általában 30–60 perc hosszúságúak és altatott patkányok agykérgéből (szürkeállományából) kerültek rögzítésre 32–384 csatornás implantátumokkal. Az adatok regisztrálásához használt implantátumok speciális, nagy elektróda sűrűségű eszközök: a beültethető szárukon található nagyszámú érzékelő szorosan egymás mellett helyezkedik el (néhány mikrométer távolságra egymástól), lehetővé téve az agyi elektromos tevékenység nagy téri felbontású mintavételezését. A nagy téri felbontás egyik legfőbb előnye, hogy egy adott idegsejt által generált elektromos impulzust több szomszédos érzékelő is észlel, és – mivel az impulzus nagysága és alakja függ az érzékelő és a neuron távolságától – ez az impulzus eltérő hullámalakokkal jelenik meg a szomszédos érzékelők jelén, az adott neuron egyfajta egyedi elektromos „lenyomatát” eredményezve (5. ábra). Ez a többcsatornás elektromos „lenyomat” segíti és megbízhatóbbá, pontosabbá teszi a neuronok azonosítását, vagyis a sejtválogatás folyamatát. A pilot projekt során archivált kutatási adatcsomagok a nyers mérések és a részletes metaadat információ mellett a sejtválogatás eredményeit, valamint a kiválogatott egyedi idegsejtek

különbéle térbeli és időbeli jellemzőit, tulajdonságait is tartalmazzák. Ilyen jellemzők például, hogy átlagosan hányszor tüzeltek a neuronok egy másodperc alatt, milyen hosszú és mekkora nagyságú az akciós potenciáljuk, vagy hogy hány csatornán észlelhető egy adott neuron elektromos impulzusa. A nyers adatok és az egyedi idegsejtek minőségére vonatkozó különféle leírók is megtalálhatók a megosztott adatcsomagban (pl. mennyire voltak tiszták/zajosak a rögzített jelek), mely segíti a megfelelő (vagy akár a legjobb) minőségű mérések, kísérletek kiválasztását.



5. ábra: Két idegsejt többcsatornás elektromos „lennyomata”. Az adatok rögzítéséhez egy 128 érzékelővel rendelkező implantátumot használtunk (az érzékelők elrendezése a bal oldalon látható). A középső és jobb oldali színes ábrákon markáns különbségek figyelhetők meg a két neuron akciós potenciál hullámformája között, mind térben, mind pedig időben (Horváth és mtsai (2021) alapján), (ms – millimásodperc, μV – mikrovolt, μm – mikrométer).

A projekt időtartama alatt három korábbi, nyílt hozzáférésű tanulmányunk (Fiáth és mtsai, 2019; Fiáth és mtsai, 2021; Horváth és mtsai, 2021) kutatási adatait archiváltuk és tettük szabadon elérhetővé a CONCORDA adatrepozitóriumban a tudományterületen használt egyik szabványos adatformátumban, metaadatokkal kiegészítve, a FAIR (megtalálható, hozzáférhető, átjárható, újrafelhasználható) alapelveknek megfelelően (Wilkinson és mtsai, 2016). A fejezet további részében röviden ismertetem a három tanulmányt és a kísérletsorozatok során keletkezett sokcsatornás mérési adatok főbb jellemzőit.

A kutatócsoport 2019-ben megjelent publikációjában (Fiáth és mtsai, 2019)³ nagy téri felbontású, szilícium-alapú agyi implantátumokkal azt vizsgáltuk, hogy az implantátum agyszövetbe történő beültetési sebessége milyen hatással van az eszközzel mért agyi elektromos tevékenység minőségére, azaz romlanak vagy javulnak-e a rögzített jelek, attól függően, hogy lassabban vagy gyorsabban juttatjuk be az implantátumot az agyszövetbe. Ehhez altatott patkányok agykérgéből rögzítettük az elektromos aktivitást egy 128 érzékelőt tartalmazó implantátummal, négy különböző beültetési sebességet alkalmazva a kísérlet során. Ily módon négy darab, 45 perc hosszúságú adatfájlunk keletkezett minden kísérletben. Ezt a protokollt összesen tíz kísérletben ismételtük meg, ugyanazt a négyféle beültetési sebességet alkalmazva, csak más-más sorrendben. A leglassabb és leggyorsabb beültetési sebességekkel további kísérleteket végeztünk, egyrészt egy másik típusú, 32 csatornás eszközzel, másrészt szövettani vizsgálatok céljából. Összesen 68 darab nagy méretű (3–13 GB), sokcsatornás adatfájlunk keletkezett a kísérletsorozatban, hozzávetőleg 0.7 TB össz mérettel.

A második tanulmányban (Fiáth és mtsai, 2021)⁴ azt vizsgáltuk, hogy van-e különbség a rögzített jelek minőségében, ha azokat sokcsatornás, nagy téri felbontású implantátumok szárának különböző részein

3 <https://science-data.hu/dataset.xhtml?persistentId=doi:10.5072/FK2/QWSIXM>

4 <https://science-data.hu/dataset.xhtml?persistentId=doi:10.5072/FK2/CGMYAH>

(szélén vagy közepén) elhelyezkedő érzékelőkkel rögzítjük. Ehhez ötféle, különböző tulajdonságokkal rendelkező, szilícium-alapú implantátummal (32–384 közötti csatornaszámmal) mértük altatott patkányok agykérgi elektromos tevékenységét. Néhány esetben más kutatócsoportok által megosztott, nyílt hozzáférésű mérési adatokat is felhasználtunk a vizsgálatokhoz. A 20–60 perces felvételek száma eszközönként eltérő volt (6–21 darab mérés, összesen 54 felvétel), és a különböző időtartamokból, valamint az eltérő csatornaszámokból adódóan ezen felvételek mérete is jelentős különbségeket mutatott (2–28 GB). A kísérletsorozatban keletkezett nyers mérési adatok összmérete nagyjából 0.8 TB-ot tett ki, ebbe beleértve a külső forrásból származó adatokat is.

A harmadik adathalmaz (Horváth és mtsai, 2021)⁵ egy olyan kutatási projektben keletkezett, ahol a fentebb már említett 128 érzékelős implantátummal rögzítettük agykérgi idegsejtek aktivitását különböző agykérgi területekről és rétegekből. A 30–60 perc hosszúságú adatfájlok mérete 8–17 GB közötti, az adatsomag összesen 109 nyers mérési adatot tartalmaz, megközelítőleg 1.2 TB összmérettel. Az adatsomagot a kutatási adatok bemutatására és megosztására specializálódott Scientific Data folyóiratban publikáltuk, és a publikáció megjelenésével egyidejűleg elérhetővé tettük a német G-Node (German Neuroinformatics Node) adatarhívumában, a GIN-ben (G-Node INfrastructure).⁶ Az ennél az adatrepozitóriumnál tapasztalt lassú letöltési sebességek miatt azonban hasznosnak és szükségesnek láttuk ezeknek az adatoknak a duplikálását egy másik adatrepozitóriumban.

Mind a három fenti adatsomagból kinyertük az egyedi idegsejtek aktivitását a sejtválogatás módszerével (ami összességében több ezer egyedi neuronból álló adathalmazt eredményezett), így a nyers mérési adatok mellett ennek az eredményét is archiváltuk és megosztottuk, a fejezet első bekezdésében felsorolt további jellemzőkkel együtt. Az adatok bemutatására, megjelenítésére az adatsomagok mellé készítettünk és

5 <https://science-data.hu/dataset.xhtml?persistentId=doi:10.5072/FK2/OKPT5U>

6 <https://doi.gin.g-node.org/10.12751/g-node.ar7ol/>

feltöltöttünk egy MATLAB programnyelv alapú programkódot (szkriptet) is, mely alapvető adatvizualizációs funkciókkal rendelkezik, így kiválóan használható a megosztott adatok megtekintésére.

4. Szabványos adatformátumok az idegtudomány területén

Az agyi elektromos tevékenység rögzítésére számtalan különféle elektrofiziológiai mérőrendszer létezik, ezek egy része kereskedelmi forgalomban kapható termék, más részük saját fejlesztésű rendszer. Ezek a mérőrendszerek általában a saját egyedi, sokszor zárt fájlformátumukban mentik el a rögzített adatokat, és gyakran ezeket a fájlformátumokat csupán a gyártó cég, általában költséges, saját fejlesztésű szoftverével lehet használni, illetve az elérhető nyílt forráskódú programok és programkódok csupán egy korlátozott része képes ezeket az adatfájlokat beolvasni. Ebből látható, hogy ha azt a célt szeretnénk elérni, hogy minél szélesebb körben felhasználják a megosztott kutatási adatainkat, törekednünk kell arra, hogy az adatokat olyan formátumban rögzítsük vagy olyan adatformátumra alakítsuk át, melyet az adatfeldolgozásra és elemzésre fejlesztett nyílt és kereskedelmi forgalomban kapható szoftverek többsége képes beolvasni, vagy az adott tudományterületen gyakran használt programnyelveken (pl. Python, MATLAB, R) írt programkódokkal könnyű kezelni.

A kutatási adatok tárolására, megosztására, használatára és feldolgozására fejlesztett szabványos adatformátumok kiválóan alkalmasak lehetnek a fenti feladatra, továbbá segíthetik a kísérleti eredmények reprodukálását, különböző laboratóriumok adatainak és eredményeinek összehasonlítását, az adatok újrafelhasználását, vagy a kutatócsoportok közötti együttműködéseket. A csillagászat területén például már több mint 40 éve kidolgozták az első szabványos adatformátumot (FITS, Flexible Image Transport System), amit azóta is széles körben használnak, rendkívüli mértékben segítve a terület kutatásait, valamint az asztronómiai adatok archiválását és megosztását. Az elektrofiziológia területén a különböző adatformátumok egységesítése, és szabványos adatformátumok kifejlesztése az elmúlt évtized

közepén kezdődött el és jelenleg is tart. A kísérletes elektrofiziológia területén két szabványos adatformátum terjedt el az utóbbi években: a NIX (Neuroscience Information eXchange; Stoewer és mtsai, 2014)⁷ és az NWB (Neurodata Without Borders; Rübel és mtsai, 2019; Rübel és mtsai, 2021; Teeters és mtsai, 2015).⁸ Előbbit inkább Európában, míg utóbbit főként az Amerikai Egyesült Államokban található laboratóriumokban használják. Mindkét adatformátum a szabványos HDF5 (Hierarchical Data Format version 5) fájlformátumot használja az adatok tárolására, és gazdagon annotálható metaadattal. Jelenleg mind a NIX, mind pedig az NWB formátumot csupán néhány tucat kutatócsoport használja, de ezek a számok folyamatosan emelkednek. Természetesen a szabványos adatformátumok elterjedését jelentősen segíti az is, ha olyan eszközöket és szoftveres környezetet fejlesztenek ki ezek mellé, melyek alkalmasak az ilyen formátumban tárolt adatok közvetlen feldolgozására és elemzésére. Végül érdemes még megemlíteni a BIDS:EEG (Brain Imaging Data Structure – Electroencephalography)⁹ szabványos adatformátumot is (Gorgolewski és mtsai, 2016; Pernet és mtsai, 2019), mely a humán EEG regisztrátumok tárolására és megosztására fejlesztett adatstruktúra.

A mi választásunk az NWB adatformátumra esett, pontosabban annak újabb verziójára az NWB:N 2.0-ra (Neurodata Without Borders: Neurophysiology version 2.0), mivel több olyan kiemelkedő, az USA-ban található élettani labor is ezt használja, melyek hasonló területen vagy kísérleti metodikával tevékenykednek, mint kutatócsoportunk (pl. a Seattle-ben található Allen Institute for Brain Science, Buzsáki György és Losonczy Attila laborjai New Yorkban, vagy Soltész Iván laboratóriuma a Stanford Egyetemen). Az NWB formátumot különféle fiziológiás módszerekkel kinyert adatok tárolására fejlesztették, többek között ide tartozik az intracellulárisan (sejten belülről) vagy extracellulárisan (sejtek közötti térből) rögzített elektromos idegi tevékenység, az optikai fiziológiás kísérletek mérései (ilyen például a kétfoton pásztázó lézermikroszkópiás módszerrel végzett kalcium képalkotás), vagy különféle kísérletes ingerlési paradigmák adatainak tárolása.

7 <http://g-node.github.io/nix/>

8 <https://www.nwb.org/>

9 <https://bids.neuroimaging.io/>

5. Hazai és nemzetközi adatrepozitóriumok idegtudományi kutatási adatok befogadására

A közlésre beadott vagy elfogadott tanulmányokhoz felhasznált kutatási adatok megosztása már több rangos tudományos folyóiratnál is kötelező (akár már a szakmai bírálati folyamat során), többek között azért, hogy könnyebben rá lehessen akadni az elemzés során elkövetett esetleges hibákra, ellenőrizni lehessen az eredmények reprodukálhatóságát, vagy, hogy ezzel megakadályozzák a kutatási adatokból kinyert eredmények sajnálatos módon néha előforduló manipulálását. A kutatási adatokat általában az adatok tárolására és megosztására szakosodott intézményi vagy nagyobb hazai és nemzetközi szervezetek által létrehozott repozitóriumokban lehet elhelyezni. Vannak általános adatrepozitóriumok, melyek bármelyik tudományterületről fogadnak adatokat, az ismertebbek közé tartozik például a figshare,¹⁰ a Zenodo¹¹ vagy az Open Science Framework (OSF).¹² Léteznek tudományterület-specifikus repozitóriumok is, az idegtudomány területén ilyenek például a G-Node,¹³ az OpenNeuro,¹⁴ az EBRAINS,¹⁵ a DANDI (Distributed Archives for Neurophysiology Data Integration)¹⁶ vagy a NeuroMorpho,¹⁷ hogy csak néhányat említsünk a legnagyobbak és legjelentősebbek közül. Ezek közül több nem csupán adatrepozitórium, hanem komplex kutatási infrastruktúra különféle kutatást segítő szolgáltatásokkal. A hazai tudományos közösség számára készült első jelentős adatrepozitórium az ELKH Számítástechnikai és Automatizálási Kutatóintézete által fejlesztett CONCORDA,¹⁸ mely 2020-ban kezdte meg működését, de az adatrepozitórium fejlesztése jelenleg is zajlik az ELKH Adatrepozitórium Projekt keretein belül.¹⁹

10 <https://figshare.com/>

11 <https://zenodo.org/>

12 <https://osf.io/>

13 <https://gin.g-node.org/>

14 <https://openneuro.org/>

15 <https://ebrains.eu/service/share-data>

16 <https://www.dandiarchive.org/>

17 <https://neuromorpho.org/>

18 <https://science-data.hu/>

19 <https://science-research-data.hu/>

A repozitóriumok között jelentős különbségek lehetnek a feltöltött adatokra vonatkozó mennyiségi- és méretkorlátokban (pl. az egyedileg feltölthető fájl méretében, az adatsomag méretében vagy a feltölthető fájlok számában), az esetleges költségekben (sok repozitórium ingyenes egy adott méretkorlátig), a felület használatának módjában vagy egyéb követelményekben. Érdemes tehát ezeket figyelembe venni a kutatási adataink számára megfelelő adattár kiválasztásakor. A kiválasztáshoz több hasznos online segédeszközt is találunk, ilyen például a re3data (REgistry of REsearch data REpositories),²⁰ melynek felületén egyszerűen kereshetünk az elérhető adattárak között, akár tudományterületenkénti, akár országonkénti bontásban. Az oldal hasznos információkat tartalmaz az adatrepozitóriumokról, például, hogy használ-e DOI-t (Digital Object Identifier) a megosztott adatsomagok azonosításához, milyen metaadat szabványt alkalmaz, vagy milyen típusú adatokat fogad. Hasonló célt szolgál a FAIRSharing²¹ weboldal is, melyen azonban a repozitóriumok mellett különféle szabványokat (pl. szabványos adatformátumokat) és kutatásiadat-kezeléshez kapcsolódó irányelveket is lehet keresni.

6. A kutatási adatok archiválása a CONCORDA adatrepozitóriumba, tanulságok és tapasztalatok

Az adatarchiválás folyamatának első lépése a három tanulmányunk nyers és származtatott kutatási adatainak összegyűjtése és rendszerezése volt. Esetünkben a származtatott adatok a sejtválogatás eredményét jelentették. Az alkalmazott kísérleti módszereinkből kifolyólag nagy adattömegekkel dolgozunk, ezért a kutatócsoportunk rendelkezik egy több terabájt kapacitású hálózati adattároló (NAS) eszközzel, mely többszörösen védve van adatvesztés ellen. Az egyes kísérletek elvégzése után közvetlenül erre az eszközre kerül archiválásra minden nyers mérési adat, így ezek összegyűjtése nem okozott problémát. A származtatott adatok összegyűjtése viszont már okozott némi fejtörést. A nyers adatok feldolgozását és elemzését ugyanis sokszor

20 <https://www.re3data.org/>

21 <https://fairsharing.org/>

különböző személyek végzik, különböző számítógépeken és más helyszíneken, és ezen vagy egyéb okokból kifolyólag a tanulmány közzététele után a származtatott adatok néha elkallódhatnak, törlésre kerülhetnek. Példának okáért tegyük fel, hogy van egy kutató, aki egy adott projekten dolgozott, majd a projekt végeztével egy másik kutatócsoportba került, és a projekt során használt számítógépét (rajta a származtatott adatokkal) pedig egy másik kutató kapta meg. Utóbbi kutató, mivel szüksége van szabad tárterületre egy másik projekthez, lementi a számítógépen található korábbi projektadatokat külső merevlemezre, ami ezután általában egy fiókba (és ezzel gyakran a feledés homályába) kerül. Néhány esetben ezt a külső tárolót egy harmadik kutató kapja meg, aki viszont már nagy valószínűséggel nem fogja tudni, hogy honnan erednek a tárolón található adatok, és azt feltételezván, hogy korábban már archiválták ezeket az adatokat, törli azokat.

A származtatott adatok hosszú távú megőrzésére egy kutatócsoportnak gyakran sajnálatos módon nincs elegendő tárterülete, így előfordulhatnak a fentíhez hasonló esetek. Különösen a több évvel korábbi kutatási projektek alatt összegyűjtött adatok lehetnek hiányosak a nem megfelelő és nem rendszeres adatarchiválás miatt. Esetünkben az adatok egy részénél a sejtválogatás után kapott adatfájlok már nem voltak fellelhetők a NAS tárolón archivált és egyéb tárolókon található adatok között. Emiatt úgy döntöttünk, hogy a korábban a sejtválogatáshoz használt szoftver egy újabb, jelentősen továbbfejlesztett verziójával ismét megcsináljuk ezt a hiányzó adatfeldolgozási lépést. Habár ennek az extra lépésnek jelentős volt az időigénye, jelentős volt a haszna is: ezáltal értékes információkhoz jutottunk mind az eredményeink reprodukálhatóságát illetően (más kutató végezte a feldolgozást egy más verziójú szoftverrel, mégis hasonló eredményre jutva), mind pedig a tanulmány eredményeinek megbízhatóságát tekintve (a tanulmány eredményeit és konklúzióját nem változtatta meg az új feldolgozás eredménye).

Mivel a projektünk esetében összességében nagy adattömegről volt szó (kb. 2.5 TB), az adatok összegyűjtését követően azokat nagy sebességű külső SSD (solid state drive) tárolókra másoltuk, hogy a szabványos adatformátumba történő átkonvertálást, a különféle jellemzők

kinyerését és az adatcsomagok mozgatását fel tudjuk gyorsítani. Mind a nyers, mind pedig a származtatott adatok, valamint az adatokat leíró és minőségüket bemutató értékek (kb. 20-30 különféle érték kiszámítása) is az NWB fájllokba kerültek becsomagolásra. Az NWB fájllok elkészítéséhez először MATLAB programnyelven elkészítettük a szükséges szkripteket, illetve módosítottuk egy korábbi projektből meglévőket. Az NWB formátumhoz Python (PyNWB)²² és MATLAB (MatNWB)²³ alapú alkalmazásprogramozási felületet (API) is fejlesztettek, melynek a segítségével, a felület megismerése után és némi programozási tapasztalattal, elkészíthetők az NWB adatfájlok. Mi a MatNWB API-t használtuk erre a feladatra. Az NWB formátum mélyebb megismerése és a MatNWB használatának elsajátítása érdekében a pilot projekt során részt vettünk egy angol nyelvű online kurzuson, de korábbi adatarchiválási és programozási tapasztalataink is segítettek a tanulási folyamatban. Érdemes itt megemlíteni az NWB fájllok egyik előnyét. Mivel a HDF5 adatstruktúra képezi az NWB fájllok alapját, ezért lehetőség van az adatok tömörítésére, akár jelentős méretcsökkenést is elérve a végleges adatállományban, ezzel pedig akár az adatok tárolásának költségeit is csökkenteni tudjuk. Tömörítés alkalmazásával az adataink eredeti összméretét mintegy ötödével, nagyjából 2 terabájtra sikerült lecsökkentenünk. A tömörítés hátránya, hogy az adatok beolvasásának ideje kis mértékben nőhet, mivel azokat előbb ki kell csomagolni.

Az NWB fájllok elkészítése után következett az adatok feltöltése a CONCORDA repozitóriumba. A Dataverse²⁴ nyílt platformon alapuló CONCORDA adatrepozitórium használata az angol nyelvű leírás alapján könnyen megtanulható. Mind új tárolót, mind pedig új adatcsomagokat nehézségek nélkül létre tudtunk hozni a felületen. Egy tárolóba készíthetünk új tárolókat vagy adatcsomagokat, és utóbbiba helyezhetjük az adatainkat. Az Integratív Idegtudományi Kutatócsoport tárolójában létrehoztunk három adatcsomagot, mind a három tanulmányunkhoz egyet-egyét. Jelenleg mind a három adatcsomag nyilvános

22 <https://pynwb.readthedocs.io/>

23 <https://github.com/NeurodataWithoutBorders/matnwb/>

24 <https://dataverse.org/>

(az elérhetőségüket lásd fentebb a 3. fejezetben). A felhasználó szempontjából a metaadatok szerkesztése (pl. szerzők, kulcsszavak, kapcsolódó közlemények, használt szoftverek vagy egyéb megjegyzések megadása) és a fájlok feltöltése is viszonylag egyszerű. A kutatóintézetten belülről az adatfájlok feltöltése és letöltése is gyorsnak bizonyult, a 10 GB feletti mérettel rendelkező adatfájlok is problémamentesen feltölthetők voltak (jelenleg 100 GB a feltöltési korlát egy fájl esetén). Véleményem szerint hasznos lenne, ha az alapértelmezett „Tábla” nézet mellett a „Fa” nézetben is lehetőségünk lenne egyszerre több fájl kiválasztására és letöltésére (pl. egy adott könyvtárban lévő összes fájl letöltése), mivel utóbbi nézet használata az adataink könyvtárszerkezete miatt praktikusabb. Megpróbálkoztunk curl és Python alapú szkripteket használva a fájlok automatikus, kötegelt feltöltésével is, mely módszer egy programozásban járatos felhasználó számára a platform részletes dokumentációjának köszönhetően viszonylag gyorsan elsajátítható. Próbaképpen kisebb fájlokat sikeresen fel is tudtunk tölteni ezzel a módszerrel. Curl alapú programkóddal nagyobb méretű (>2GB) fájlok feltöltése is lehetséges, azonban mivel összességében viszonylag kevés, de nagy méretű feltöltendő adatfájlunk volt (a nyers és a származtatott adatok is a szabványos fájlformátumokba kerültek becsomagolásra), így végül a webes grafikus felhasználói felületről (GUI) elérhető egyszerű és kényelmes fájlfeltöltési módszer mellett döntöttünk.

Először a G-Node (GIN) nemzetközi adatrepozitóriumban elérhető kutatási adataink duplikálását végeztük el a CONCODA-ba, ezzel jelentősen javítva az adataink elérhetőségét. A körülbelül 0.9 TB összméretű és 112 fájl tartalmazó adatcsomagot néhány óra alatt fel tudtuk tölteni a webes felületet használva. Ezt követte a két másik adatcsomag feltöltése hasonló módon. Minden adatcsomaghoz készítettünk egy rövid angol nyelvű leírást a fájlok elnevezésének logikájáról, a könyvtárszerkezet hierarchiájáról, a szabványos fájlformátum struktúrájáról, valamint rövid példakódokat is mellékelünk a fájlok beolvasására MATLAB környezetben, illetve egy Python-alapú célszoftverben. Továbbá minden adatcsomaghoz feltöltöttünk két CSV (Comma Separated Values) fájlt, melyek a kísérletekhez és kísérleti állatokhoz kapcsolódó jellemzőket

(pl. állat súlya, neme, megcélzott agyi terület), valamint a mérési adatokhoz kapcsolódó információkat (pl. fájl mérete, hossza, minősége) tartalmazták minden mérési adatfájlhoz.

Végül az adatok beolvasásához és megjelenítéséhez készítettünk egy MATLAB-alapú programkódot, melyet minden adatcsomag mellé feltöltöttünk az adatrepozitóriumba. A kód példákat mutat be a nyers mérési adatok, valamint a kinyert idegsejtek és tulajdonságaik beolvasására és megjelenítésére, továbbá alapszintű jelfeldolgozási lépéseket is (pl. jelek szűrése) tartalmaz. A szkriptet dokumentáltuk is, így a programkód könnyen átalakítható úgy, hogy különböző típusú agyi implantátumok adatait, vagy a mérési adatok különböző részeit is tanulmányozni lehessen vele. Az archivált adatcsomagok bibliográfiai leírását végül a Magyar Tudományos Művek Tárába (MTMT) is felvittük kutatási adattípusként.

Hivatkozásjegyzék

- Fiáth R, Meszéna D, Somogyvári Z, Boda M, Barthó P, Ruther P, Ulbert I. Recording site placement on planar silicon-based probes affects signal quality in acute neuronal recordings. (2021) SCIENTIFIC REPORTS 11(1): 2028. <https://doi.org/10.1038/s41598-021-81127-5>
- Fiáth R, Márton AL, Mátyás F, Pinke D, Márton G, Tóth K, Ulbert I. Slow insertion of silicon probes improves the quality of acute neuronal recordings. (2019) SCIENTIFIC REPORTS 9 (1): 111. <https://doi.org/10.1038/s41598-018-36816-z>
- Fiáth R, Raducanu BC, Musa S, Andrei A, Lopez CM, van Hoof C, Ruther P, Aarts A, Horváth D, Ulbert I. A silicon-based neural probe with densely-packed low-impedance titanium nitride micro-electrodes for ultrahigh-resolution in vivo recordings. (2018) BIOSENSORS & BIOELECTRONICS 106: 86–92. <https://doi.org/10.1016/j.bios.2018.01.060>

- Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, Flandin G, Ghosh SS, Glatard T, Halchenko YO, Handwerker DA, Hanke M, Keator D, Li X, Michael Z, Maumet C, Nichols BN, Nichols TE, Pellman J, Poline JB, Rokem A, Schaefer G, Sochat V, Triplett W, Turner JA, Varoquaux G, Poldrack RA. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. (2016) *SCIENTIFIC DATA* 3(1): 160044. <https://doi.org/10.1038/sdata.2016.44>
- Horváth C, Tóth LF, Ulbert I, Fiáth R. Dataset of cortical activity recorded with high spatial resolution from anesthetized rat (2021) *SCIENTIFIC DATA* 8(1): 180. <https://doi.org/10.1038/s41597-021-00970-3>
- Pernet CR, Appelhoff S, Gorgolewski KJ, Flandin G, Phillips C, Delorme A, Oostenveld R. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. (2019) *SCIENTIFIC DATA* 6(1): 103. <https://doi.org/10.1038/s41597-019-0104-8>
- Raducanu BC, Yazıcıoğlu RF, Lopez CM, Ballini M, Putzeys J, Wang S, Andrei A, Rochus V, Welkenhuysen M, Van Helleputte N, Musa S, Puers R, Kloosterman F, Van Hoof C, Fiáth R, Ulbert I, Mitra S. Time multiplexed active neural probe with 1356 parallel recording sites. (2017) *SENSORS* 17(10): 2388. <https://doi.org/10.3390/s17102388>
- Rübel O, Tritt AJ, Ly R, Dichter BK, Ghosh SS, Niu L, Soltesz I, Svoboda K, Frank LM, Bouchard K. The Neurodata Without Borders ecosystem for neurophysiological data science. (2022) *BIORXIV* 435173 <https://doi.org/10.1101/2021.03.13.435173>
- Rübel O, Tritt A, Dichter B, Braun T, Cain N, Clack N, Davidson TJ, Dougherty M, Fillion-Robin JC, Graddis N, Grauer M, Kiggins JT, Niu L, Ozturk D, Schroeder W, Soltesz I, Sommer FT, Svoboda K, Ng L, Frank LM, Bouchard K. NWB:N 2.0: An Accessible Data Standard for Neurophysiology. (2019) *BIORXIV* 523035 <https://doi.org/10.1101/523035>

- Steinmetz NA, Aydın C, Lebedeva A, Okun M, Pachitariu M, Bauza M, Beau M, Bhagat J, Böhm C, Broux M, Chen S, Colonell J, Gardner RJ, Karsh B, Kloosterman F, Kostadinov D, Lopez CM, O'Callaghan J, Park J, Putzeys J, Sauerbrei B, van Daal RJJ, Vollan AZ, Wang S, Welkenhuysen M, Ye Z, Dudman JT, Dutta B, Hantman AW, Harris KD, Lee AK, Moser EI, O'Keefe J, Renart A, Svoboda K, Häusser M, Haesler S, Carandini M, Harris TD. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. (2021) *SCIENCE* 372(6539): eabf4588. <https://doi.org/10.1126/science.abf4588>
- Stoewer A, Kellner CJ, Benda J, Wachtler T and Grewe J. File format and library for neuroscience data and metadata. (2014) *FRONTIERS IN NEUROINFORMATICS CONFERENCE ABSTRACTS: NEUROINFORMATICS* <https://doi.org/10.3389/conf.fninf.2014.18.00027>
- Teeters, JL, Godfrey K, Young R, Dang C, Friedsam C, Wark B, Asari H, Peron S, Li N, Peyrache A, Denisov, G, Siegle JH, Olsen SR, Martin C, Chun M, Tripathy S, Blanche TJ, Harris K, Buzsáki G, Koch C, Meister M, Svoboda K, Sommer FT Neurodata without borders: creating a common data format for neurophysiology. (2015) *NEURON* 88(4): 629–634. <https://doi.org/10.1016/j.neuron.2015.10.025>
- Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship. (2016) *SCIENTIFIC DATA* 3(1): 160018. <https://doi.org/10.1038/sdata.2016.18>