

Správa výzkumných dat

Open science: od dat k publikaci



CENTRUM PRO PODPORU
OPEN SCIENCE
Univerzita Karlova

Dagmar Hanzlíková & Milan Janíček

Projekt RKV II

3. 2. 2021, v 10:00

Centrum pro podporu open science

openscience@cuni.cz



Tato prezentace byla podpořena projektem OP VVV Rozvoj kapacit pro výzkum a vývoj UK II, CZ.02.2.69/0.0/0.0/18_054/0015222



UNIVERZITA
KARLOVA



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



HR EXCELLENCE IN RESEARCH



Open Science



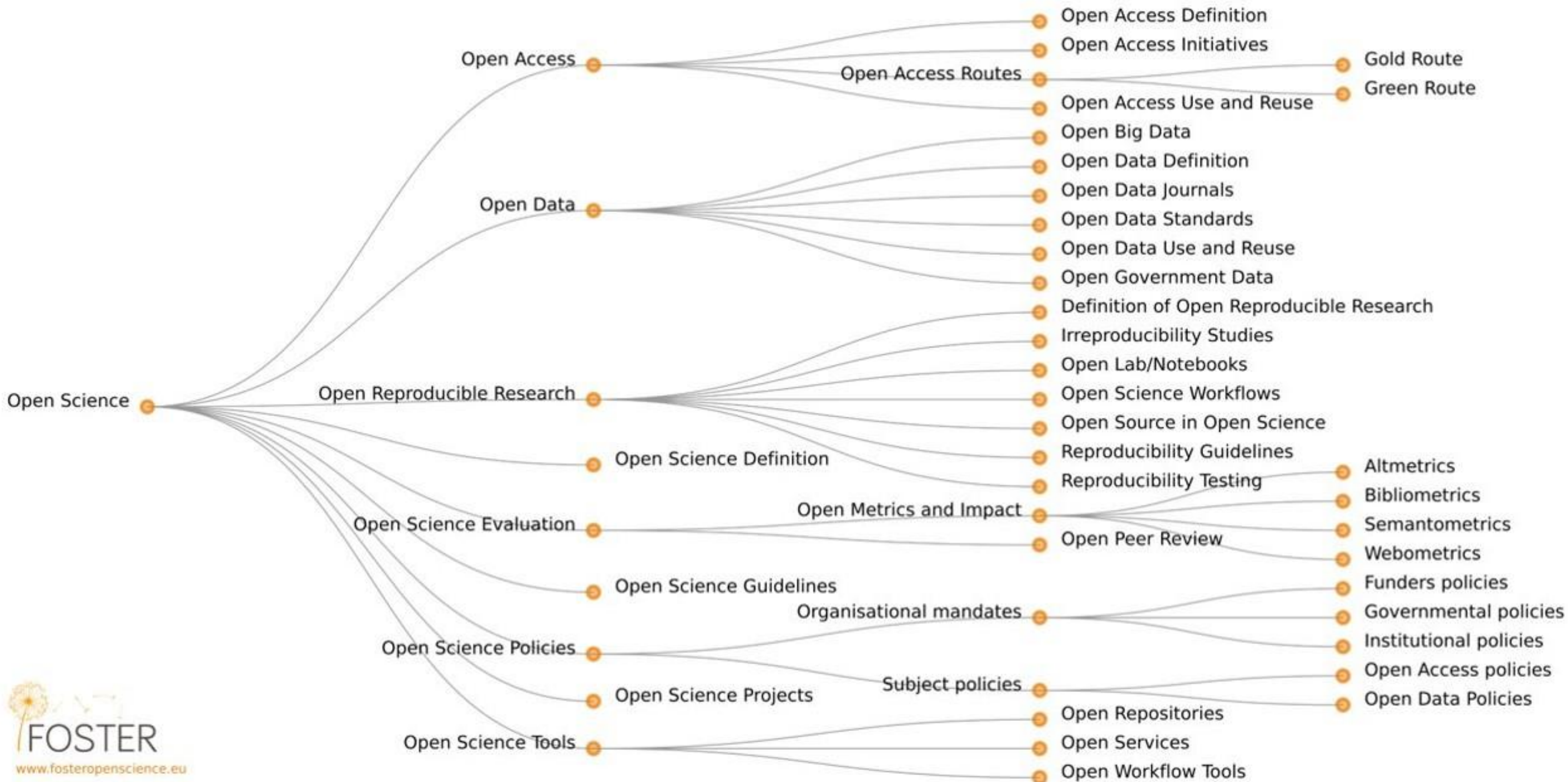
Open Science Definition

- **Open Science** is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.

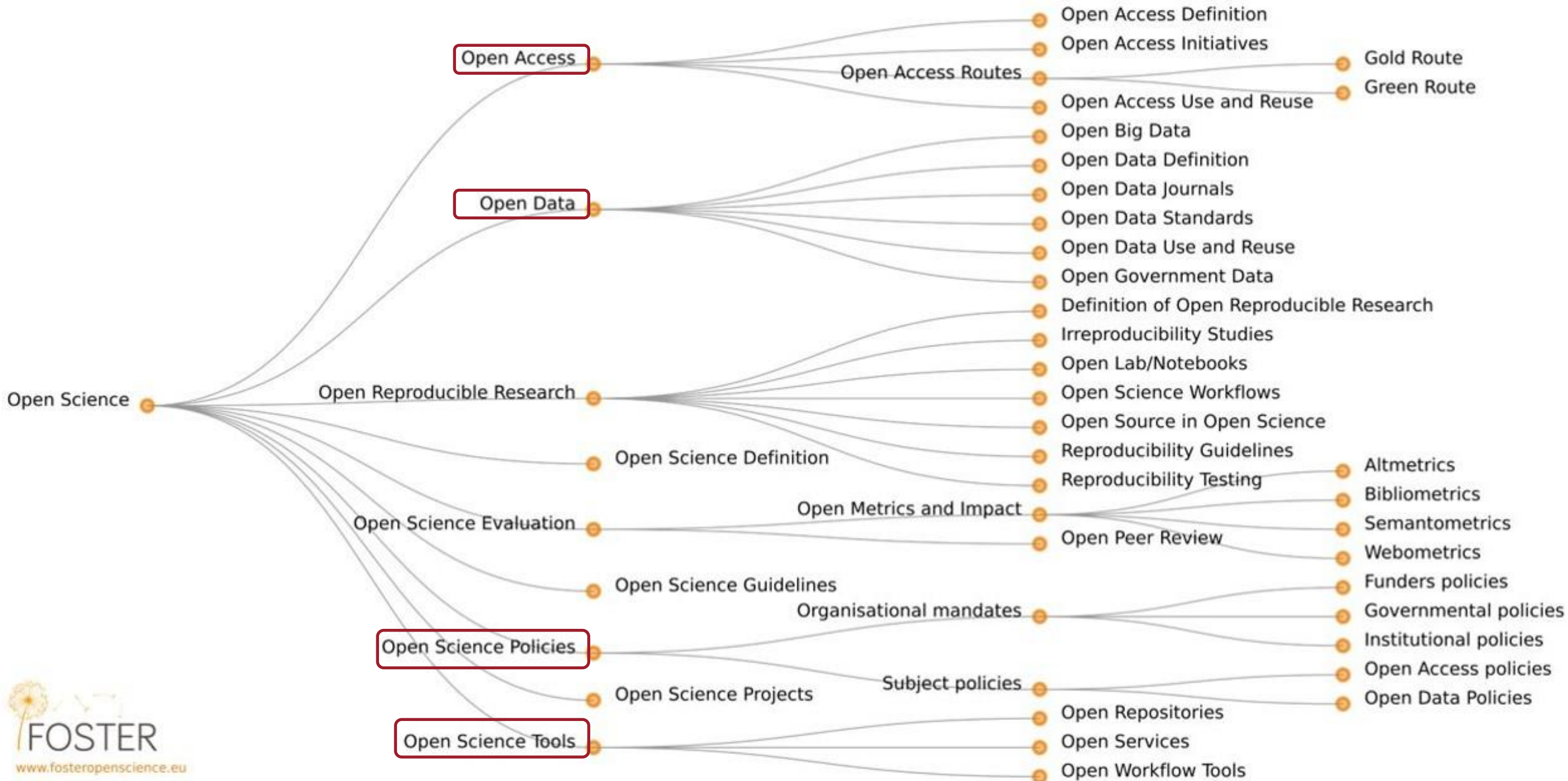
<https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>



Open Science Taxonomy

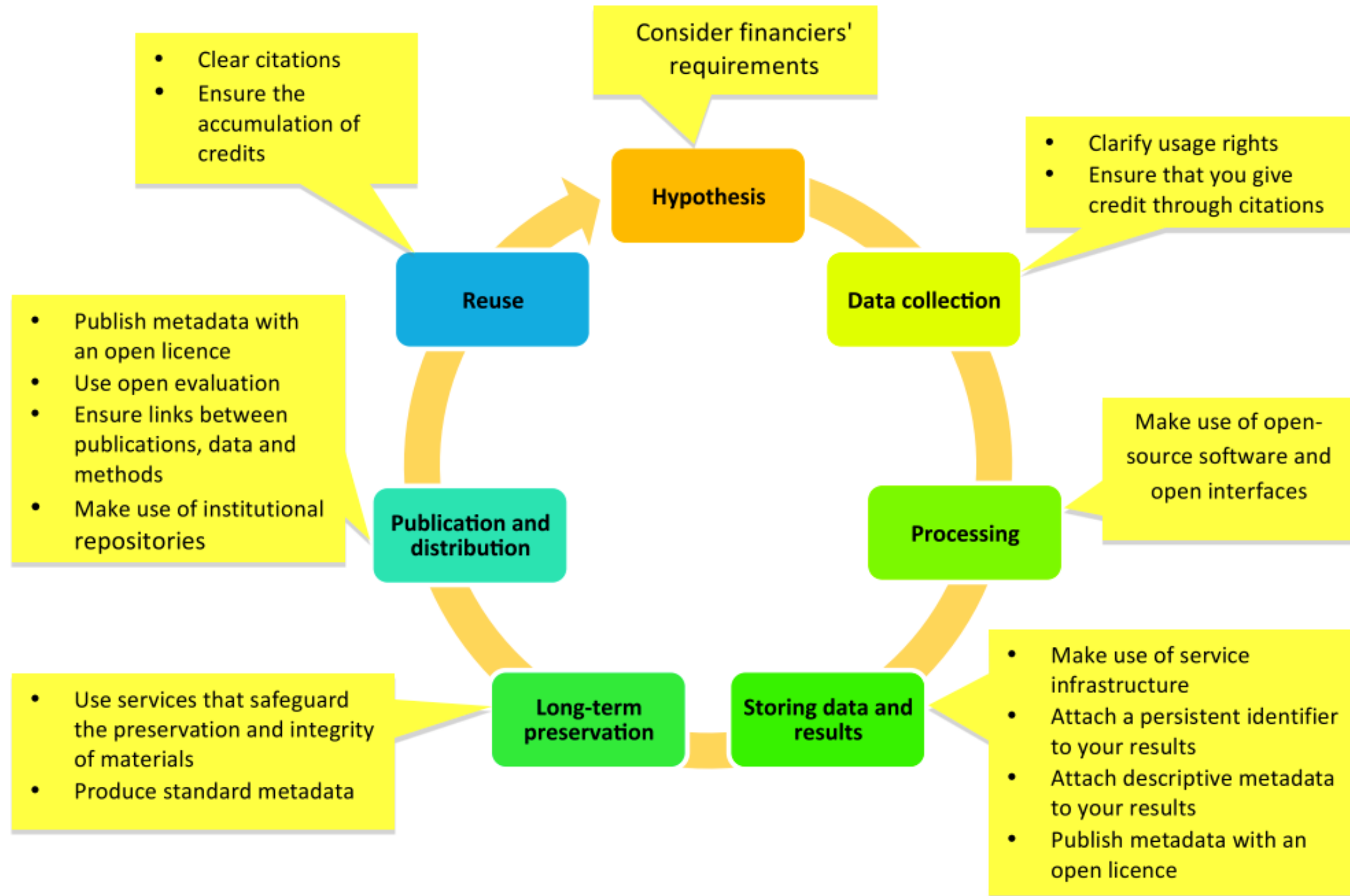


Open Science Taxonomy



Vliv Open Science na výzkumný proces

- Open Science není jen o otevřeném publikování výstupů – má vliv na různé fáze výzkumného procesu
 - formulování hypotézy
 - sběr dat
 - zpracování dat
 - ukládání dat a výsledků
 - dlouhodobé uchování
 - publikování a distribuce výsledků
 - znovupoužití



Open Science a data

- s technologickým rozvojem se významně rozšířily možnosti práce s daty
- internet umožňuje sdílet zdrojová data k publikacím - jak to udělat správně?
- v rámci Open Science přichází 2 nejpodstatnější doporučení:
 - jaká **kritéria** by měla publikovaná data splňovat
 - na to, jak vybrat **která** data zveřejňovat a za jakých podmínek



2 principy

1. jak by měla data vypadat => **FAIR principy**

- Findable Accessible Interoperable Reusable

2. která data zveřejňovat => **“as open as possible, as closed as necessary”**

- ne všechna data lze zcela volně zveřejnit – je třeba brát v potaz ochranu osobních údajů, intelektuální vlastnictví, další oprávněné zájmy...
- FAIR principy a „aoapacan“ jsou v současnosti obecně považovány za základní principy pro sdílení vědeckých dat a požadavky na jejich dodržování jsou postupně implementovány i do legislativy v rámci EU

Jak vytvářet data odpovídající novým požadavkům?

- Pokud mají být data sdílena (ať už úplně volně, nebo v nějaké omezené formě) je potřeba s nimi správně pracovat...

=> roste význam **správy výzkumných dat** a jejího **plánování**





Správa výzkumných dat



Co jsou výzkumná data?

= informace, které byly vytvořeny za účelem zjištění či reprodukování vašich výsledků výzkumu

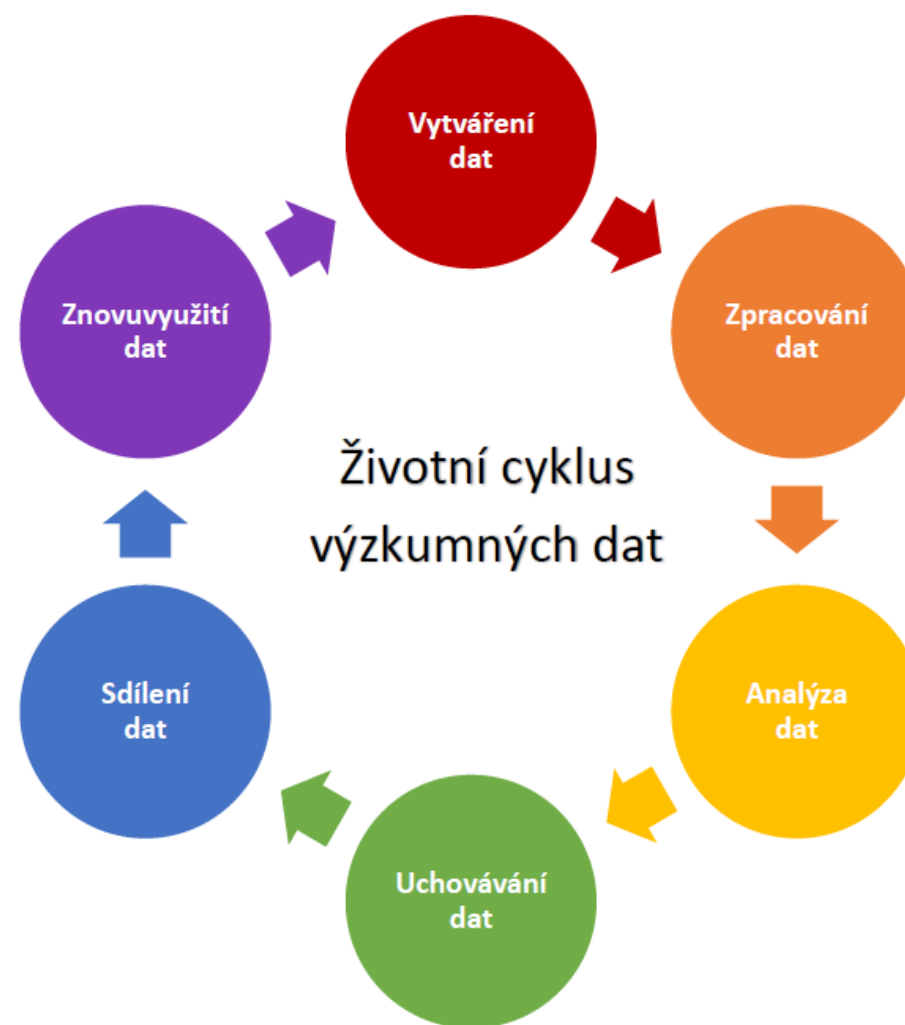
- Výzkumná data mohou mít různou podobu (digitální i nedigitální)
 - Tabulky, dokumenty
 - Audio a video nahrávky, obrázky, fotografie
 - Dotazníky, odpovědi na otázky, přepisy rozhovorů
 - Laboratorní deníky, terénní poznámky
 - Software, skript
 - Vzorčky, exempláře



Správa výzkumných dat

= organizování, skladování a dlouhodobé uchovávání

- Snazší psaní a revize článků
- Lepší zabezpečení dat, nižší riziko jejich ztráty
- Kontinuita dlouhodobých projektů
- Integrita výzkumu, lepší reputace
- Pokrok v celosvětovém výzkumu díky možnosti sdílení a opakovaného využití dat



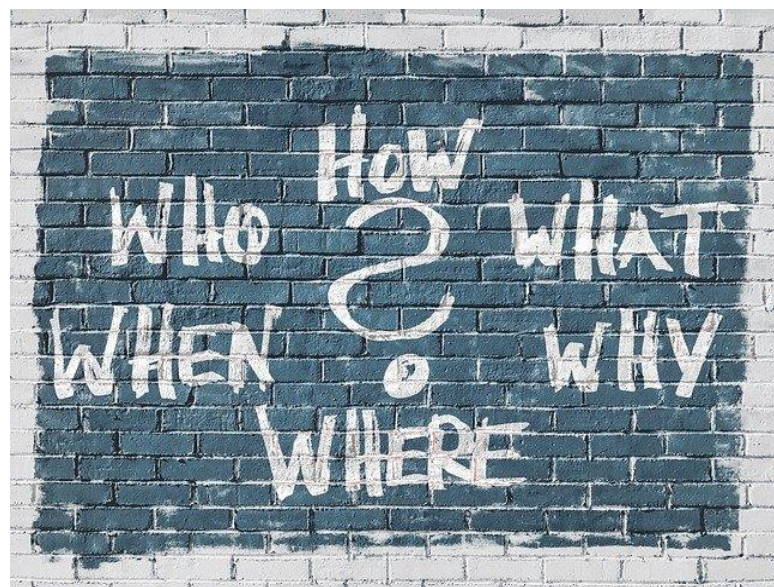


Data Management Plan



Data management plan (DMP)

= dokument, který specifikuje, *jaká data a jakým způsobem budou během výzkumu vytvářena*, a obsahuje informace o jejich *dostupnosti a využití*



Obsah DMP

1. Administrativní data
2. Sběr dat
3. Dokumentace a metadata
4. Etické a právní otázky
5. Skladování a zálohování
6. Dlouhodobá ochrana
7. Sdílení dat
8. Zodpovědnost a zdroje

Proč vytvářet DMP?

- Splnění podmínek **poskytovatelů financí** (např. H2020)
- Pomáhá předvídat **potenciální problémy**
- Snížení rizika **duplicitní práce, ztráty dat a narušení bezpečnosti**
- Zajištění, že data jsou **přesná, úplná a spolehlivá**
- Zajištění **kontinuity** a **konzistence** dlouhodobých projektů
- Pomoc při **sdílení dat**
- Ušetření času a energie (např. při znovuvyhledávání dat, obnově)

• version B: only changing the article, eg. OIP_a-WN1 → OIP_the-WN1

↓ sentence ↓ speaker ↓ article ↓ white noise → context 1

DMP: Administrativní údaje

Uved'te *základní informace*, které poskytnou kontext k danému DMP

- Základní informace o výzkumu, např. název projektu, jméno, kontaktní údaje, ID hlavního řešitele, informaci o poskytovateli financí
- Krátký popis výzkumu, ke kterému se data vztahují (abstrakt)
- Nařízení, opatření nebo směrnice, které upravují uchovávání dat na instituci (pokud existují)

DMP: Sběr dat

Definujte, *jaká data* budou vytvářena a *jakým způsobem*

- Existují data, která lze znovu využít?
- Jaké standardy nebo metodiky budete k vytváření dat používat?
- V jakém formátu budou data uchovávána?
- Jak budou složky a soubory strukturovány a pojmenovány?
- Zajistíte kontrolu kvality dat?

DMP: Dokumentace a metadata

*Poskytněte informace, které jsou potřeba pro **čtení a interpretaci dat** v budoucnosti*

- Bude vytvořena doplňující příloha s metadaty a dokumentací?
- Jak bude tato příloha vytvořena?
- Jaké metadatové standardy budou využity a proč?



DMP: Etické a právní otázky

Zvažte etické a právní záležitosti

- Máte souhlas pro uchovávání a sdílení dat?
- Jak budete chránit osobní údaje účastníků výzkumu, pokud to bude třeba?
- Budou data sdílena, nebo bude sdílení omezeno?
- Budou data licencována?



DMP: Skladování a zálohování

*Zvažte, kde budou data **skladována** a jak budou **zálohována**, včetně **přístupu** k nim a **zajištění bezpečnosti***

- Máte dostatek místa pro ukládání dat, nebo budete muset započítat náklady na úložné prostory?
- Kdo bude zodpovědný za zálohování a obnovení dat?
- Jaké je potenciální riziko pro zabezpečení dat a jak bude řešeno?
- Jak zajistíte, že se spoluřešitelé dostanou bezpečně k datům?

DMP: Dlouhodobá ochrana

*Určete, která data jsou vhodná pro **dlouhodobou ochranu** a jak nejlépe je **uchovat***

- Jaká data musí být uchována nebo zničena ze smluvních/právních důvodů?
- Jak vybrat data pro dlouhodobé uchování?
- Vyčíslili jste časové a finanční náklady na přípravu dat pro dlouhodobé uchování a sdílení?



DMP: Sdílení dat

Rozmyslete, *která data budete sdílet a jak*, a také jak vám uživatelé vašich dat budou moci *přiznat autorství*

- S kým budete data sdílet a za jakých podmínek?
- Jakým způsobem budete data sdílet?
- Kdy budou data dostupná?
- Je potřeba nastavit nějaká omezení ve sdílení dat?
- Jak se o datech dozví vaši potenciální uživatelé?



DMP: Zodpovědnost a zdroje

Určete *role a odpovědnost* pro všechny aktivity s daty v rámci výzkumu a zvažte případné *náklady s vytvářením a uchováváním dat*

- Kdo bude odpovědný za vedení DMP?
- Jak bude odpovědnost rozdělena mezi řešitele projektu?
- Potřebujete nějaké další vybavení nebo expertizu?
- Potřebujete nějaké další finanční zdroje pro sběr, zpracování a uchování dat?

Data Management Plan: Příklad



DMP: Administrativní údaje - příklad

- název:
 - A role of striatal M4 muscarinic acetylcholine receptors in learning of stereotypic motor sequences
- řešitel : Helena Janíčková PhD. , FGÚ AV ČR
- abstrakt :
 - Striatum is the main input nucleus of the basal ganglia, important for the learning, emitting and updating of stereotypic motor sequences (habits). While the M4 muscarinic acetylcholine receptors are abundantly expressed in the striatum, their role in acquiring and changing habits is not known. Therefore, in this pilot study, we use M4-preferential antagonist tropicamide to test how blocking these receptors affects learning of a complex motor sequence in mice. During the study, mice are trained to emit multiple presses of the left and right lever in the operant box in the predefined order and times. The precise sequence and time constraints of the presses are regularly changed during the training. In selected sessions, a group of mice is systemically injected with tropicamide to determine its effect on emitting of a well-learned motor sequence, on the ability to learn a new one and the ability to adjust it when the contingencies change.

DMP: Sběr dat - příklad

- data:
 - „master sheet“ popisující průběh pokusu (trvá několik měsíců)
 - která myš je v které části experimentu, kdy provádí session
 - program k řízení operantního boxu
 - definuje sekvenci, která je od myši požadována v jednotlivých částech tréninku (operantní box má vlastní programovací jazyk, Med Associates)
 - logy z boxů
 - záznam činnosti myši v boxu (timestamp + operace, např. správná páčka/špatná páčka/získání odměny...)
 - skripty pro analýzu konkrétních otázek a jejich výstupy
 - skripty v pythonu které vytáhnou z logů odpovědi na konkrétní analytické otázky


```
\*****
\
\          MAIN PROGRAM
\*****
S.S.2,
S1, \ Turn HouseLight and associated stimulus ON
  #START: ON ^HouseLight; ON ^LeftLight ---> S2

S2, \ press left [L]
  #R^LeftLever: OFF ^LeftLight; Z^Z_Correct1;---> S3 \to low time state
  #R^RightLever: Z^Z_IncorrectR; ---> S6 \ to bad state (turn on right light for debugging)
  #Z^Z_End:      ---> S15

\low time state 2nd press
S3,
  #R^LeftLever: Z^Z_IncorrectLLT, ---> S6 \ to bad state
  #R^RightLever: Z^Z_IncorrectR ---> S6 \ to bad state
  A(^TimeoutLowLeftTicks)#T: ON ^LeftLight ---> S4 \ to best time state
  #Z^Z_End:      ---> S15

\best time state 2nd press
S4,
  #R^LeftLever: Z^Z_Correct2; OFF ^LeftLight; OFF ^RightLight ---> S23 \ to reinforcement state [ change !!!]
  #R^RightLever: Z^Z_IncorrectR ---> S6
  A(^TimeoutHighLeftTicks)#T: OFF ^LeftLight ---> S5 \ to high time state [was error in comment]
  #Z^Z_End:      ---> S15

\high time state 2nd press
S5,
  #R^LeftLever: Z^Z_IncorrectLHT ---> S6 \ to bad state
  #R^RightLever: Z^Z_IncorrectR ---> S6 \ to bad state
  #Z^Z_End:      ---> S15

S6, \ turn off lights - Bad state 1
  0.01": OFF ^HouseLight, ^LeftLight, ^RightLight ---> S7
  #Z^Z_End:      ---> S15

S7, \turn on lights after 5" - Bad state 2
```



```
c6m1,200512,104958,0,0, 1, "", 0.000, 40.000, 0.000, 0.000, 237.000, 5.000, 0.000, 0.000, 0.000,
0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 1753.000, 1.000, 0.000, 0.000, 0.000, 75.000,
0.000, 0.000, 1.000, 0.000, 1.000, 0.050, 0.000, 60.000, 2.000, 1.000, 40.000, 1.000, 5.000, 0.000, 0.500, 3600.000, 50.000, 360000.000, 0.000, 0.000, 0.000, 0.00
0, 0.000, 1.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.510, 1334.650, 0.500, 865.100, 1075.650, 0.500, 304.690, 0.500, 385.100, 609.650, 0.500, 213.690, 0.500, 229.690, 0.
500, 267.650, 0.500, 6.690, 0.500, 606.100, 257.100, 0.200, 192.100, 3011.650, 0.500, 667.650, 0.500, 958.100, 891.100, 0.200, 10.100, 88.100, 0.200, 1962.650, 0.500, 15
09.100, 1534.100, 0.200, 624.650, 0.500, 271.690, 0.500, 442.650, 0.500, 183.690, 0.500, 691.650, 0.500, 1527.100, 904.650, 0.500, 454.680, 0.500, 1360.650, 0.500, 264.
690, 0.500, 343.650, 0.500, 67.690, 0.500, 155.690, 0.500, 556.650, 0.500, 55.690, 0.500, 1098.650, 0.500, 272.680, 0.500, 558.650, 0.500, 1077.100, 1056.100, 0.200, 563
.650, 0.500, 1554.100, 475.100, 0.200, 5275.100, 471.100, 0.200, 2123.100, 538.100, 0.200, 2313.100, 973.100, 0.200, 1243.650, 0.500, 353.690, 0.500, 597.100, 72.100, 0
.200, 2921.650, 0.500, 382.690, 0.500, 715.650, 0.500, 9.690, 0.500, 841.100, 344.100, 0.200, 1249.100, 581.100, 0.200, 1827.100, 440.100, 0.200, 5548.100, 325.100, 0.2
00, 678.100, 929.100, 0.200, 4183.650, 0.500, 1558.100, 1404.100, 0.200, 1555.100, 233.100, 0.200, 1558.100, 588.100, 0.200, 853.100, 1233.100, 0.200, 1028.100, 355.10
0, 0.200, 2313.100, 1127.100, 0.200, 1509.650, 0.500, 2085.650, 0.500, 423.690, 0.500, 199.650, 0.500, 1279.100, 1957.100, 0.200, 7018.100, 764.650, 0.500, 2241.650, 0.
500, 398.680, 0.500, 3219.100, 1873.100, 0.200, 3151.100, 872.100, 0.200, 3003.100, 1968.100, 0.200, 1865.100, 9.660, 0.500, 2791.650, 0.500, 1802.650, 0.500, 2306.100
, 1008.650, 0.500, 736.650, 0.500, 252.690, 0.500, 1061.100, 316.100, 0.200, 6829.100, 759.100, 0.200, 3333.650, 0.500, 969.650, 0.500, 3786.100, 287.100, 0.200, 275.10
0, 16.660, 0.500, 1054.100, 206.100, 0.200, 2118.100, 1211.100, 0.200, 2699.100, 279.100, 0.200, 3853.100, 167.100, 0.200, 2519.650, 0.500, 1955.100, 234.100, 0.200, 12
52.100, 495.100, 0.200, 2752.100, 2226.100, 0.200, 1041.100, 729.100, 0.200, 1918.100, 525.100, 0.200, 2564.100, 897.100, 0.200, 1245.100, 265.100, 0.200, 63.100, 2417
.100, 0.200, 1.300, C(237), C(238), C(239), C(240), C(241), C(242), C(243), C(244), C(245), C(246), C(247), C(248), C(249), C(250), C(251), C(252), C(253), C(254), C(255
), C(256), C(257), C(258), C(259), C(260), C(261), C(262), C(263), C(264), C(265), C(266), C(267), C(268), C(269), C(270), C(271), C(272), C(273), C(274), C(275), C(276)
, C(277), C(278), C(279), C(280), C(281), C(282), C(283), C(284), C(285), C(286), C(287), C(288), C(289), C(290), C(291), C(292), C(293), C(294), C(295), C(296), C(297),
C(298), C(299), C(300), C(301), C(302), C(303), C(304), C(305), C(306), C(307), C(308), C(309), C(310), C(311), C(312), C(313), C(314), C(315), C(316), C(317), C(318), C
(319), C(320), C(321), C(322), C(323), C(324), C(325), C(326), C(327), C(328), C(329), C(330), C(331), C(332), C(333), C(334), C(335), C(336), C(337), C(338), C(339), C(
340), C(341), C(342), C(343), C(344), C(345), C(346), C(347), C(348), C(349), C(350), C(351), C(352), C(353), C(354), C(355), C(356), C(357), C(358), C(359), C(360), C(3
61), C(362), C(363), C(364), C(365), C(366), C(367), C(368), C(369), C(370), C(371), C(372), C(373), C(374), C(375), C(376), C(377), C(378), C(379), C(380), C(381), C(38
2), C(383), C(384), C(385), C(386), C(387), C(388), C(389), C(390), C(391), C(392), C(393), C(394), C(395), C(396), C(397), C(398), C(399), C(400), C(401), C(402), C(403
), C(404), C(405), C(406), C(407), C(408), C(409), C(410), C(411), C(412), C(413), C(414), C(415), C(416), C(417), C(418), C(419), C(420), C(421), C(422), C(423), C(424)
, C(425), C(426), C(427), C(428), C(429), C(430), C(431), C(432), C(433), C(434), C(435), C(436), C(437), C(438), C(439), C(440), C(441), C(442), C(443), C(444), C(445),
C(446), C(447), C(448), C(449), C(450), C(451), C(452), C(453), C(454), C(455), C(456), C(457), C(458), C(459), C(460), C(461), C(462), C(463), C(464), C(465), C(466), C
(467), C(468), C(469), C(470), C(471), C(472), C(473), C(474), C(475), C(476), C(477), C(478), C(479), C(480), C(481), C(482), C(483), C(484), C(485), C(486), C(487), C(
488), C(489), C(490), C(491), C(492), C(493), C(494), C(495), C(496), C(497), C(498), C(499), C(500), C(501), C(502), C(503), C(504), C(505), C(506), C(507), C(508), C(5
09), C(510), C(511), C(512), C(513), C(514), C(515), C(516), C(517), C(518), C(519), C(520), C(521), C(522), C(523), C(524), C(525), C(526), C(527), C(528), C(529), C(53
0), C(531), C(532), C(533), C(534), C(535), C(536), C(537), C(538), C(539), C(540), C(541), C(542), C(543), C(544), C(545), C(546), C(547), C(548), C(549), C(550), C(551
), C(552), C(553), C(554), C(555), C(556), C(557), C(558), C(559), C(560), C(561), C(562), C(563), C(564), C(565), C(566), C(567), C(568), C(569), C(570), C(571), C(572)
, C(573), C(574), C(575), C(576), C(577), C(578), C(579), C(580), C(581), C(582), C(583), C(584), C(585), C(586), C(587), C(588), C(589), C(590), C(591), C(592), C(593),
C(594), C(595), C(596), C(597), C(598), C(599), C(600), C(601), C(602), C(603), C(604), C(605), C(606), C(607), C(608), C(609), C(610), C(611), C(612), C(613), C(614), C
(615), C(616), C(617), C(618), C(619), C(620), C(621), C(622), C(623), C(624), C(625), C(626), C(627), C(628), C(629), C(630), C(631), C(632), C(633), C(634), C(635), C(
636), C(637), C(638), C(639), C(640), C(641), C(642), C(643), C(644), C(645), C(646), C(647), C(648), C(649), C(650), C(651), C(652), C(653), C(654), C(655), C(656), C(6
57), C(658), C(659), C(660), C(661), C(662), C(663), C(664), C(665), C(666), C(667), C(668), C(669), C(670), C(671), C(672), C(673), C(674), C(675), C(676), C(677), C(67
8), C(679), C(680), C(681), C(682), C(683), C(684), C(685), C(686), C(687), C(688), C(689), C(690), C(691), C(692), C(693), C(694), C(695), C(696), C(697), C(698), C(699)
```

```
#"cas mezi odmenou a zmacknutim jakekoli packy, tj. interval mezi .20 a .10 nebo .50"
def analyze_postreinforcement_LR():
    results = {}
    for mid in my_dataset.keys():
        results[mid] = {}
        for d in sorted(my_dataset[mid].keys()):
            results[mid][d] = {}
            reinf_timestamp = -1
            trial_counter = 1;
            result_set = my_dataset[mid][d]

            for t in result_set.keys():
                print ("mid"+mid+ "time: "+str(t)+" value:"+str(result_set[t]))
                if "2" in result_set[t]:
                    reinf_timestamp = t
                    trial_counter = trial_counter + 1
                elif (("1" in result_set[t]) or ("5" in result_set[t])) & (reinf_timestamp != -1):
                    results[mid][d][trial_counter]=t-reinf_timestamp
                    reinf_timestamp = -1

    print(json.dumps(results))

#output
with open(output_post_reinforcement_LR, "w", newline='') as csvfile:
    csvwriter = csv.writer(csvfile, delimiter=',')

    csvwriter.writerow(["#Postreinforcement Any (Correct or Incorrect)"])
    csvwriter.writerow(["#mouse id", "group", "day", "trial number", "time (10ms)", "time (human readable)"])
    csvwriter.writerow(["mid", "grp", "day", "trial", "time_10ms", "time"])

#group
for g in mice_order:
    #mouse
    for m in mice_order[g]:
        #day
        for d in results[m].keys():
            for tr in results[m][d].keys():
```



```
#Postreinforcement Any (Correct or Incorrect)
#mouse id,group,day,trial number,time (10ms),time (human readable)
mid,grp,day,trial,time_10ms,time
c32m3,saline,1,2,755,0m:7s
c32m3,saline,1,3,815,0m:8s
c32m3,saline,1,4,712,0m:7s
c32m3,saline,1,5,1380,0m:13s
c32m3,saline,1,6,727,0m:7s
c32m3,saline,1,7,449,0m:4s
c32m3,saline,1,8,1075,0m:10s
c32m3,saline,1,9,778,0m:7s
c32m3,saline,1,10,578,0m:5s
c32m3,saline,1,11,496,0m:4s
c32m3,saline,1,12,451,0m:4s
c32m3,saline,1,13,487,0m:4s
c32m3,saline,1,14,23,0m:0s
c32m3,saline,1,15,231,0m:2s
c32m3,saline,1,16,305,0m:3s
c32m3,saline,1,17,907,0m:9s
c32m3,saline,1,18,33,0m:0s
c32m3,saline,1,19,7951,1m:19s
c32m3,saline,1,20,285,0m:2s
c32m3,saline,1,21,2410,0m:24s
c32m3,saline,1,22,2654,0m:26s
c32m3,saline,1,23,1104,0m:11s
c32m3,saline,1,24,1930,0m:19s
c32m3,saline,1,25,1719,0m:17s
c30m2,saline,1,2,5678,0m:56s
c30m2,saline,1,3,891,0m:8s
c30m2,saline,1,4,873,0m:8s
c30m2,saline,1,5,9679,1m:36s
c30m2,saline,1,6,5531,0m:55s
c30m2,saline,1,7,2495,0m:24s
c30m2,saline,1,8,2158,0m:21s
c30m2,saline,1,9,3184,0m:31s
c30m2,saline,1,10,354,0m:3s
c30m2,saline,1,11,11854,1m:58s
"output_post_reinforcement_LR.csv" [dos] 262L, 8671C
```

DMP: Dokumentace a metadata - příklad

- podrobný popis metodiky bude součástí publikace
- program boxů
 - kód vychází ze specifikace programovacího jazyka
 - nově vytvořený kód je důsledně komentovaný
- logy z boxů
 - kódy operací jsou standardní pro boxy + nově přidáné jsou popsány ve zdrojovém kódu programů boxů
- metadata budou vytvořena na základě požadavků repozitáře ve kterém budou data uložena



DMP: Etické a právní otázky

- etika
 - experiment na zvířatech podléhá schválení etickou komisí a odpovídá požadavkům direktivy EU
 - Directive of the European Community Council on the use of laboratory animals (2010/63/EU)
- sdílení dat není omezeno
- software bude zveřejněn pod otevřenou licencí
 - (firma vyrábějící boxy by chtěla vytvořit repozitář kódů)



DMP: Skladování a zálohování - příklad

- velikost dat i skriptů je zanedbatelná (max MB)
- data jsou uchovávána na sdílené pracovní stanici v FGÚ, probíhá standardní zálohování
- sdílení dat probíhá prostřednictvím přímého přístupu k pracovní stanici

DMP: Dlouhodobá ochrana - příklad

- není nastaven proces pro dlouhodobé uchování dat
 - cena za uchování dat by neměla být vzhledem k jejich charakteru příliš velká
- bude řešeno v budoucnu, součástí řešení může být uložení do institucionálního repozitáře AV ČR



DMP: Sdílení dat - příklad

- data budou sdílena společně se softwarem
- software – zejména programy pro boxy může být sdílen ještě dalším kanálem (např. využít specializovaný repozitář)
- data budou dostupná poté, co budou publikována



DMP: Zodpovědnost a zdroje - příklad

- zodpovědnosti
 - hlavní řešitel
 - student provádějící pokusy
 - programování softwaru (boxy a analýza)
- v tuto chvíli není potřeba dalších finančních zdrojů



Nástroje pro tvorbu DMP



Procedure

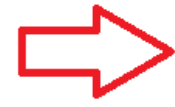
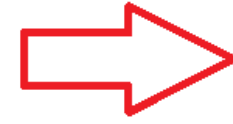
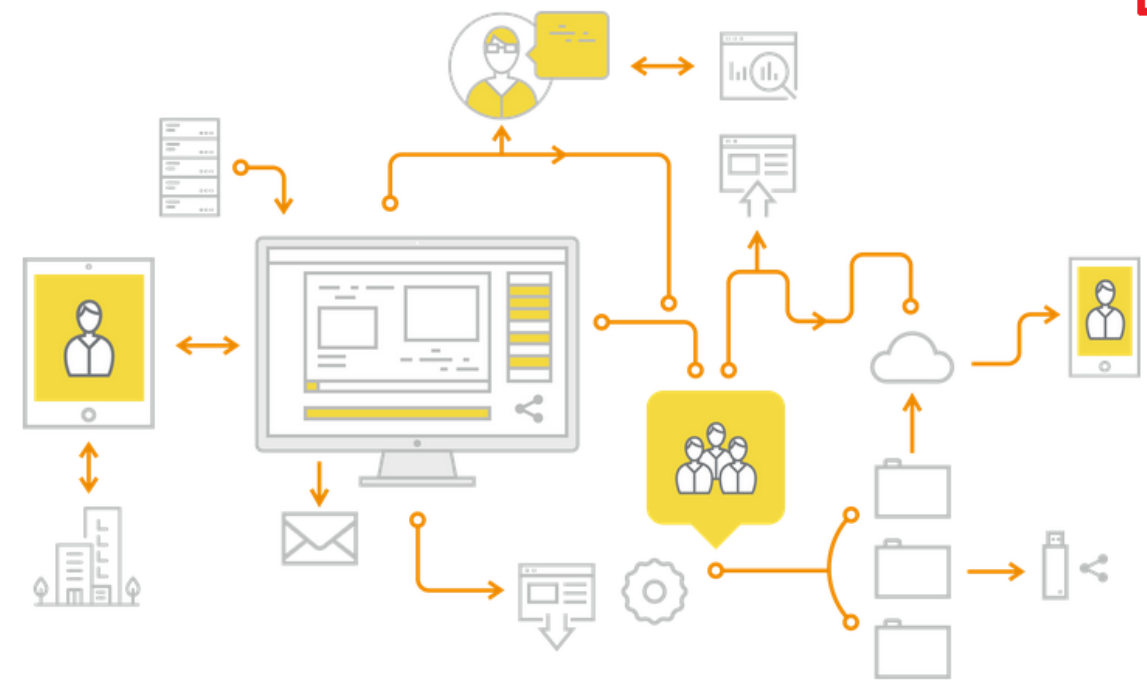
- 2 native speakers of English (Claire and Pete) recorded a list of sentences (20 pairs) in which both definite and indefinite article would be ~~both~~ contextually possible (although one of them possibly more likely) - for the list of sentences, see the Phonemic restoration folder
 - normalize the intensity of the recordings to 70dB, so that some of them wouldn't be louder than others
 - label the recordings in Praat - all boundaries at zero crossings
 - Tier 1: only the article, in the case of zero article one phoneme on each side of the gap
 - Tier 2: article + 1 adjacent phoneme on each side, for 0 article same as Tier 1
 - Tier 3: article + 2 adjacent phonemes on each side, for 0 article two phonemes on each side of the gap
 - Tier 4: both words on the sides (just to see the whole phrase)
 - duration check (statistical analysis)
 - run a script over tiers 1-3, measure the duration of the middle interval (where the article is)
 - compare durations of intervals with "a" and "the" / zero and "the"
 - t-test, paired, one tail (because we expect "the" to be longer)
 - 1) zero articles included: $p = 0.08$ for tier 1, $p < 0.001$ for tiers 2-3
 - 2) only "a" and "the": $p < 0.05$ for all three tiers
- ⇒ there is a statistically significant difference between intervals with definite and indefinite articles, although mind that comparing 40 sentences or less (20 sentences by 2 speakers) is not a great statistical power

on labelling:

Machac, Skarnitzl (2009):
Principles of Phonetic
Segmentation
(guidelines)

Plan to make data work for you

Data Management Plans that meet institutional funder requirements.



Sign in **Create account**

* **Email**

* **Password**

[Forgot password?](#)

Remember email

Sign in

- or -

Sign in with your institutional credentials

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

dmponline.dcc.ac.uk





Create a new plan

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

* What research project are you planning?

mock project for testing, practice, or educational purposes

* Select the primary research organisation

- or - No research organisation associated with this plan or my research organisation is not listed

* Select the primary funding organisation

- or - No funder associated with this plan or my funder is not listed

Create plan

Cancel

A role of striatal M4 muscarinic acetylcholine receptors in learning of stereotypic motor sequences

Project Details

Plan overview

Write Plan

Share

Download

* Project title

A role of striatal M4 muscarinic acetylcholine receptors in learni

mock project for testing, practice, or educational purposes

Funder

Grant number

e.g. 123456

Project abstract

Striatum is the main input nucleus of the basal ganglia, important for the learning, emitting and updating of stereotypic motor sequences (habits). While the M4 muscarinic acetylcholine receptors are abundantly expressed in the striatum, their role in acquiring and changing habits is not known. Therefore, in this pilot study, we use M4-preferential antagonist tropicamide to test

ID

70551

Principal Investigator

Select Guidance

To help you write your plan, DMPonline can show you guidance from a variety of organisations.

Select up to 6 organisations to see their guidance.

- Digital Curation Centre
- FAIRsFAIR - Fostering Fair Data Practices in Europe

Find guidance from additional organisations below

[See the full list](#)

Save

Set plan visibility



Public or organisational visibility is intended for finished plans. You must answer at least 50% of the questions to enable these options. Note: test plans are set to private visibility by default.

- Private: visible to me, specified collaborators and administrators at my organisation
- Organisation: anyone at my organisation can view
- Public: anyone can view



Manage collaborators

Invite specific people to read, edit, or administer your plan. Invitees will receive an email notification that they have access to this plan.

Email address	Permissions
	Owner

Invite collaborators

* Email



* Permissions

- Co-owner
- Editor
- Read only

Co-owner: Has admin-rights to the plan (can invite other users, view the plan, answer questions, or comment). Editor: Has edit-rights to the plan (can view the plan, answer questions, or comment). Read Only: Has read-rights to the plan (can view the plan or comment)

Submit

A role of striatal M4 muscarinic acetylcholine receptors in learning of stereotypic motor sequences

Project Details Plan overview **Write Plan** Share Download

expand all | collapse all 13/13 answered

Data Collection (2 / 2)

What data will you collect or create?

B *I* [List] [Table] [Link] [Grid]

Following kinds of data will be created:

- „master sheet“
 - contains the information about the experiment - which mouse is in which part of the experiment, dates of specific sessions
 - master sheet will be updated manually by researchers
 - master sheet will be stored in XSL (MS Excel) format
- operant conditioning box program
 - defines sequence of operations that is requested in respective parts of mice's training
 - the program will be written in device specific programming language (MedAssociates)
- operant conditioning box logs
 - logs are stored automatically by the operant box
 - all mouse's interactions with operant box are recorded into log as two values
 - timestamp
 - operation code (eg. correct press, wrong press, issuing of reinforcement)
 - the data are in plain text structure of the data (semantics) is described in

Guidance	Comments
DCC	FAIRsFAIR

Questions to consider:

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

Guidance:

Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of

A role of striatal M4 muscarinic acetylcholine receptors in learning of stereotypic motor sequences

Project Details Plan overview Write Plan Share **Download**

Download settings

Optional Plan Components

- project details coversheet
- question text and section headings
- unanswered questions

Format

pdf

PDF formatting

Font

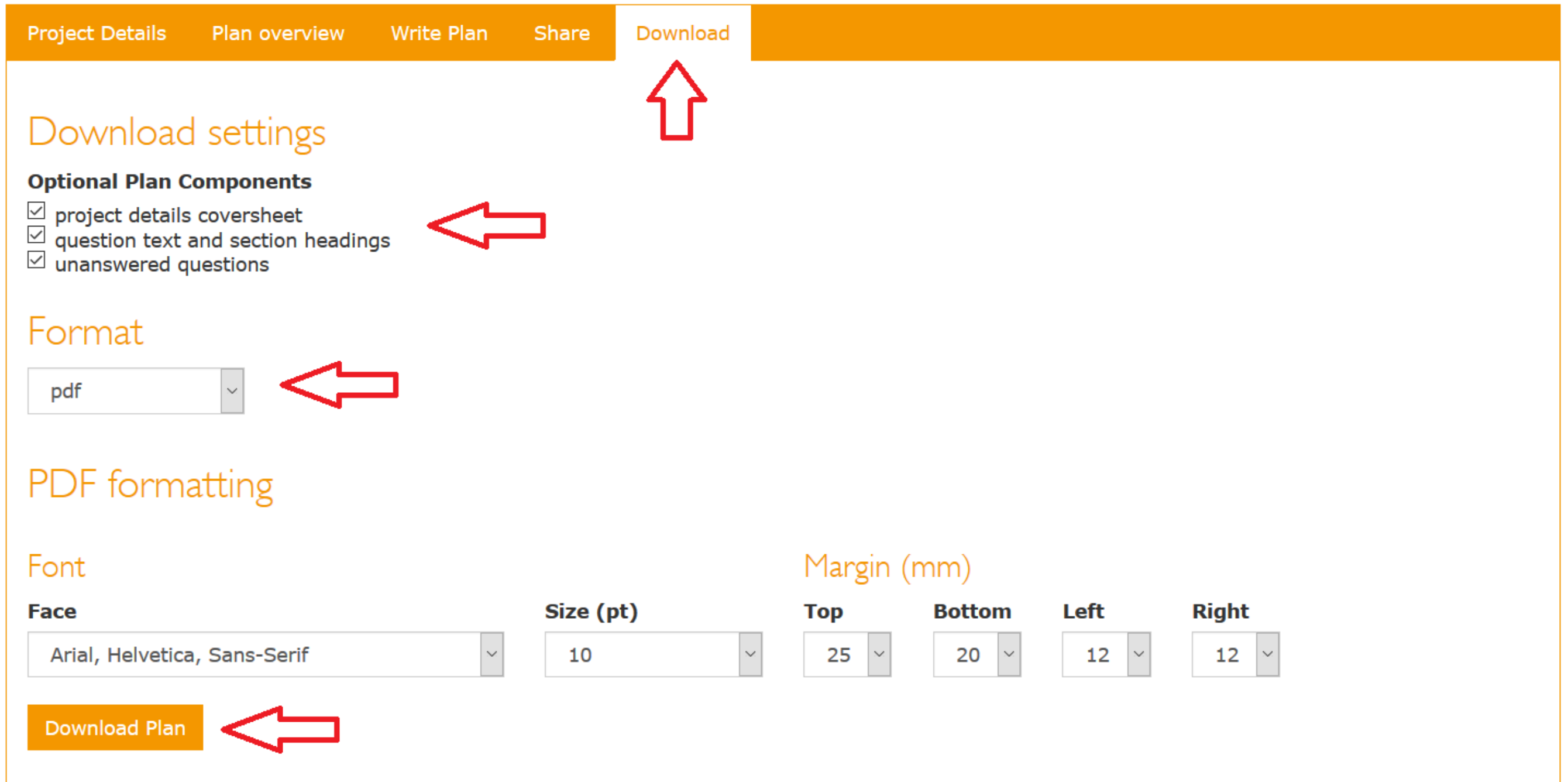
Face: Arial, Helvetica, Sans-Serif

Size (pt): 10

Margin (mm)

Top	Bottom	Left	Right
25	20	12	12

Download Plan



A role of striatal M4 muscarinic acetylcholine receptors in learning of stereotypic motor sequences

A Data Management Plan created using DMPonline

Creator: Milan Janíček

Affiliation: Univerzita Karlova

Template: DCC Template

Project abstract:

Striatum is the main input nucleus of the basal ganglia, important for the learning, emitting and updating of stereotypic motor sequences (habits). While the M4 muscarinic acetylcholine receptors are abundantly expressed in the striatum, their role in acquiring and changing habits is not known. Therefore, in this pilot study, we use M4-preferential antagonist tropicamide to test how blocking these receptors affects learning of a complex motor sequence in mice. During the study, mice are trained to emit multiple presses of the left and right lever in the operant box in the predefined order and times. The precise sequence and time constraints of the presses are regularly changed during the training. In selected sessions, a group of mice is systemically injected with tropicamide to determine its effect on emitting of a well-learned motor sequence, on the ability to learn a new one and the ability to adjust it when the contingencies change.

Last modified: 31-01-2021

A role of striatal M4 muscarinic acetylcholine receptors in learning of stereotypic motor sequences

Data Collection

What data will you collect or create?

Following kinds of data will be created:

- „master sheet“
 - contains the information about the experiment - which mouse is in which part of the experiment, dates of specific sessions
 - master sheet will be updated manually by researchers
 - master sheet will be stored in XSL (MS Excel) format
- operant box program
 - defines sequence of operations that is requested in respective parts of mice's training
 - the program will be written in device specific programming language (MedAssociates)
- operant box logs
 - logs are stored automatically by the operant box
 - all mouse's interactions with operant box are recorded into log as two values
 - timestamp
 - operation code (eg. correct press, wrong press, issuing of reinforcement)
 - the data are in plain text, structure of the data (semantics) is described in MedAssociates documentation
 - operation codes are dependant on operant conditioning box program
- scripts for specific analysis and their outputs
 - scripts which take specific information from the logs (eg. time length of trial for a mouse) and store them in output file
 - scripts are written in Python
 - outputs are stored in CSV files

How will the data be collected or created?

Experiment data will be produced by operant box running selected programs. Further data will be created by scripts based on original operant box logs.

Documentation and Metadata

What documentation and metadata will accompany the data?

Complete methodics will be part of published paper.
Description of codes generated by the operant box will be provided in separate file.
All programs and scripts are well commented.

Ethics and Legal Compliance

How will you manage any ethical issues?

The project was approved by ethical commission and fulfils requirements stated in Directive of the European Community Council on the use of laboratory animals (2010/63/EU)

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Sharing of produced data will not be limited. Source codes will be published under an open source license.

Storage and Backup

How will the data be stored and backed up during the research?

Size of the data shall not reach more than Megabytes.
Data are stored on shared work station in the lab. Data backups are created regularly and according to Institute's internal policies.

How will you manage access and security?

Access and security are handled with respect to Institute's policy.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Master sheet, operant box programs and experiment logs should be preserved long-term to ensure reproducibility of results.

What is the long-term preservation plan for the dataset?

TBD

Data Sharing

How will you share the data?

Data will be provided as part of published paper. Data will be shared in a data repository. Source codes will be also shared in source code repositories.

Are any restrictions on data sharing required?

All data will be shared after the paper is published.

Responsibilities and Resources

Who will be responsible for data management?

Principal investigator is responsible for overall data management.
Experiments are run by students working on the project.
Programs and scripts are created and updated by external collaborator.

What resources will you require to deliver your plan?

There are no additional resources needed.



Podpora na UK



DMP skupina

Podpora vědců při **tvorbě DMP** a **práci s výzkumnými daty**

➤ Přednáška o podpoře na UK v **17:00**

dmp@cuni.cz





Prostor pro diskuzi



Co nás dnes ještě čeká?

Přestávka na oběd 11:30 – 12:30		
Nástroje pro správu výzkumných dat a oborové repozitáře	12:30 – 14:00	Mgr. Jan Urban, Ph.D. Ing. Marek Suchánek Mgr. Jindřich Krejčí, Ph.D. Mgr. Pavel Straňák, Ph.D.
Přestávka 14:00 – 14:15		
Sdílení výzkumných dat: As open as possible, as closed as necessary	14:15 – 15:15	Mgr. Dagmar Hanzlíková Ing. Milan Janíček
EU quō vādis? Evropské vědecké infrastruktury a open science	15:15 – 16:15	Ing. Milan Janíček
Přestávka 16:15 – 16:30		
Služby e-infrastruktury CESNET	16:30 – 17:00	RNDr. David Antoš, Ph.D.
Podpora na UK: Co nám kdy univerzita dala?	17:00 – 18:00	Mgr. Dagmar Hanzlíková Ing. Milan Janíček

Užitečné odkazy

- [DMPonline](#): Nástroj pro tvorbu DMP
- [FAIRsharing](#): Nástroj pro vyhledání oborových metadatových standardů a databází
- [Re3data.org](#): Registr repozitářů výzkumných dat
- [Projekt GDPR na UK](#): Informace o GDPR na UK, legislativa, návody, vzory
- [DCC How-to Guides](#): Návody a checklisty ke správě výzkumných dat



Použité obrázky

- FOSTER Open Science Taxonomy [online]. Dostupné z https://www.fosteropenscience.eu/themes/fosterstrap/images/taxonomies/os_taxonomy.png Podléhá licenci CC-BY
- [xx] Clker-Free-Vector-Images. Mouse. *Pixabay* [online]. Dostupné z <https://pixabay.com/images/id-312012/> Podléhá licenci [Pixabay](#).
- [xx] Geralt. questions-qui-quoi-comment. *Pixabay* [online]. Dostupné z <https://pixabay.com/images/id-2110967/> Podléhá licenci [Pixabay](#).
- [21] Screenshot <https://dmponline.dcc.ac.uk/> 31. 1. 2021





Děkujeme za pozornost.

Centrum pro podporu open science
Ústřední knihovna UK
openscience@cuni.cz

