

# TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models

**Rastko Ciric**

Department of Psychology, Stanford University

**William Thompson**

Department of Clinical Neuroscience, Karolinska Institutet

**Romy Lorenz**

MRC CBU, University of Cambridge

**Mathias Goncalves**

Massachusetts Institute of Technology

**Eilidh MacNicol**

Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London

**Christopher Markiewicz**

Department of Psychology, Stanford University

**Yaroslav Halchenko**

Department of Psychological and Brain Sciences, Dartmouth College

**Satrajit Ghosh**

Massachusetts Institute of Technology <https://orcid.org/0000-0002-5312-6729>

**Krzysztof Gorgolewski**

Stanford University <https://orcid.org/0000-0003-3321-7583>

**Russell Poldrack**

Stanford <https://orcid.org/0000-0001-6755-0259>

**Oscar Esteban** (✉ [phd@oscaresteban.es](mailto:phd@oscaresteban.es))

CHUV University of Lausanne <https://orcid.org/0000-0001-8435-6191>

---

## Resource

**Keywords:** TemplateFlow, brain models, atlases, neuroimaging

**Posted Date:** August 26th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-264855/v3>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)



# TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models

Rastko Ciric<sup>1,2\*</sup>, William H. Thompson<sup>3</sup>, Romy Lorenz<sup>1,4,5</sup>, Mathias Goncalves<sup>1</sup>, Eilidh MacNicol<sup>6</sup>, Christopher J. Markiewicz<sup>1</sup>, Yaroslav O. Halchenko<sup>7</sup>, Satrajit S. Ghosh<sup>8,9</sup>, Krzysztof J. Gorgolewski<sup>1</sup>, Russell A. Poldrack<sup>1</sup>, Oscar Esteban<sup>10\*</sup>

\*For correspondence:  
rastko@stanford.edu (RC);  
phd@oscaresteban.es (OE)

<sup>1</sup>Department of Psychology, Stanford University, CA, USA; <sup>2</sup>Department of Bioengineering, Stanford University, CA, USA; <sup>3</sup>Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; <sup>4</sup>MRC CBU, University of Cambridge, Cambridge, UK; <sup>5</sup>Department of Neurophysics, MPI, Leipzig, Germany; <sup>6</sup>Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; <sup>7</sup>Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA; <sup>8</sup>McGovern Institute for Brain Research, Massachusetts Institute of Technology: MIT, Cambridge, MA, USA; <sup>9</sup>Department of Otolaryngology, Harvard Medical School, Boston, MA, USA; <sup>10</sup>Department of Radiology, University Hospital of Lausanne and University of Lausanne, Lausanne, Switzerland

---

**Abstract** Reference anatomies of the brain and corresponding atlases play a central role in experimental neuroimaging workflows and are the foundation for reporting standardized results. The choice of such references —i.e., templates— and atlases is one relevant source of methodological variability across studies, which has recently been brought to attention as an important challenge to reproducibility in neuroscience. *TemplateFlow* is a publicly available framework for human and nonhuman brain models. The framework combines an open database with software for access, management, and vetting, allowing scientists to distribute their resources under FAIR —findable, accessible, interoperable, reusable— principles. *TemplateFlow* supports a multifaceted insight into brains across species, and enables multiverse analyses testing whether results generalize across standard references, scales, and in the long term, species, thereby contributing to increasing the reliability of neuroimaging results.

---

1 Brains are morphologically variable, exhibiting diversity in such features as overall size (*Lüders*  
2 *et al.*, 2002), sulcal curvature (*Tosun et al.*, 2015), and functional topology (*Tavor et al.*, 2016; *Mars*  
3 *et al.*, 2018). Morphological variability manifests not only in differences between brains but also  
4 in the way that a brain changes across its lifespan, as it is remodelled by development, aging, and  
5 degenerative processes (*Courchesne et al.*, 2000; *Good et al.*, 2001; *Sowell et al.*, 2003). These mor-  
6 phological differences often correspond with the effects of interest in neuroimaging studies and  
7 hinder direct spatial comparisons between brain maps (*Brett et al.*, 2002). The substantial vari-  
8 ability within and between individual brains necessitates a means of formalizing population-level  
9 knowledge about brain anatomy and function. Neuroscientists have answered this need by cre-  
10 ating brain atlases as references for understanding and contextualizing morphological variability.  
11 Atlases map landmarks, features, and other knowledge about the brain as annotations that are  
12 consistent across individual brains.

13 The development of atlases in neuroscience has accelerated knowledge discovery and dissem-  
14 ination. Early endeavors, epitomized by the groundbreaking work of *Brodmann (2006)*, originally

15 published in German in 1909) and complemented by *Von Economo and Koskinas (2008*, originally  
16 published in German in 1925), leveraged careful scrutiny of microanatomy and cytoarchitectonic  
17 properties in small numbers of brains. Concurrent macroanatomical approaches, by contrast, iden-  
18 tified common features in nuclear boundaries and cortical gyrification. Modern atlases advanced  
19 on these approaches by incorporating stereotaxy, defining a basis set of coordinate axes over  
20 the brain and anchoring neural landmarks to coordinates. Talairach's assiduous postmortem ex-  
21 amination of a single brain produced a stereotaxic atlas that saw wide use (*Talairach et al., 1957*).  
22 Stereotaxy was a fundamental feature to unfold surgical neuronavigation systems. *Schurr and Mer-*  
23 *rington (1978)* developed the first stereotaxic apparatus to surgically induce targeted brain lesions  
24 on cats. This early antecedent of neuronavigation informed early sectional atlases of the cat and  
25 macaque brains. Since then, neuroscientists have directed great efforts to improve existing (*Ta-*  
26 *lairach and Tournoux, 1988*) and generate new atlases of the neurotypical adult human (*Landman*  
27 *et al., 2012*) and nonhuman (*Paxinos and Watson, 1997; Martin and Bowden, 2000*) brain; as well  
28 as developing, aging, and neurologically atypical brains. For instance, new atlases and represen-  
29 tative stereotaxic maps can be created for diseased (*Dickie et al., 2015*), infant (*Matsuzawa et al.,*  
30 *2001; Fonov et al., 2011; Shi et al., 2011*), and elderly (*Buckner et al., 2004*) human populations or  
31 to capture the rapid postnatal development of nonhuman species (*Calabrese et al., 2013; Szulc*  
32 *et al., 2015*). Advancing beyond the volumetric constraints of stereotaxy, researchers of primate  
33 neocortex have also devised standard spaces based on geometric reconstructions of the cortical  
34 surface. This surface-based approach has the advantage of respecting the intrinsic topology of  
35 cortical folds, a development that has led to further improvements in spatial localization (*Coalson*  
36 *et al., 2018*). On account of its relatively high spatial resolution, its capacity to image the entire  
37 brain, and its non-invasive acquisition protocols, magnetic resonance imaging (MRI) has revolu-  
38 tionized neuroscience in general and the atlasing endeavor (*Evans et al., 2012*) in particular. In  
39 combination with software instruments' progress to map homologous features between subjects  
40 supported by regular grids (*Avants et al., 2008*) or reconstructed anatomical surfaces (*Robinson*  
41 *et al., 2014*), MRI has enabled researchers to create population-average maps of a particular image  
42 modality and/or particular sample with relative ease. These maps, called "templates", are typically  
43 created by averaging features across individuals that are representative of the population of inter-  
44 est to a study (*Dickie et al., 2017*). As a result, atlasing endeavours have been made contingent on  
45 templates, and have largely shifted away from the search for a single universal neuroanatomical  
46 pattern, instead making use of increasingly large samples with the aim of representing a popula-  
47 tion average of the distribution of morphological patterns.

48 Such resources as atlases and templates, which provide standardized prior knowledge, have be-  
49 come an indispensable component of modern neuroimaging data workflows for two cardinal rea-  
50 sons. First, group inference in neuroimaging studies requires that individuals' features are aligned  
51 into a common spatial frame of reference where their location can be called standard (*Brett et al.,*  
52 *2002*). Second, templates engender a stereotaxic coordinate system in which atlases can be de-  
53 linedated or projected. Associating atlases with template coordinates also facilitates the mapping  
54 of prior population-level knowledge about the brain into images of individual subjects' brains (for  
55 instance, to sample and average the functional MRI signal indexed by the regions defined in an  
56 atlas; *Yeo et al. (2011)*).

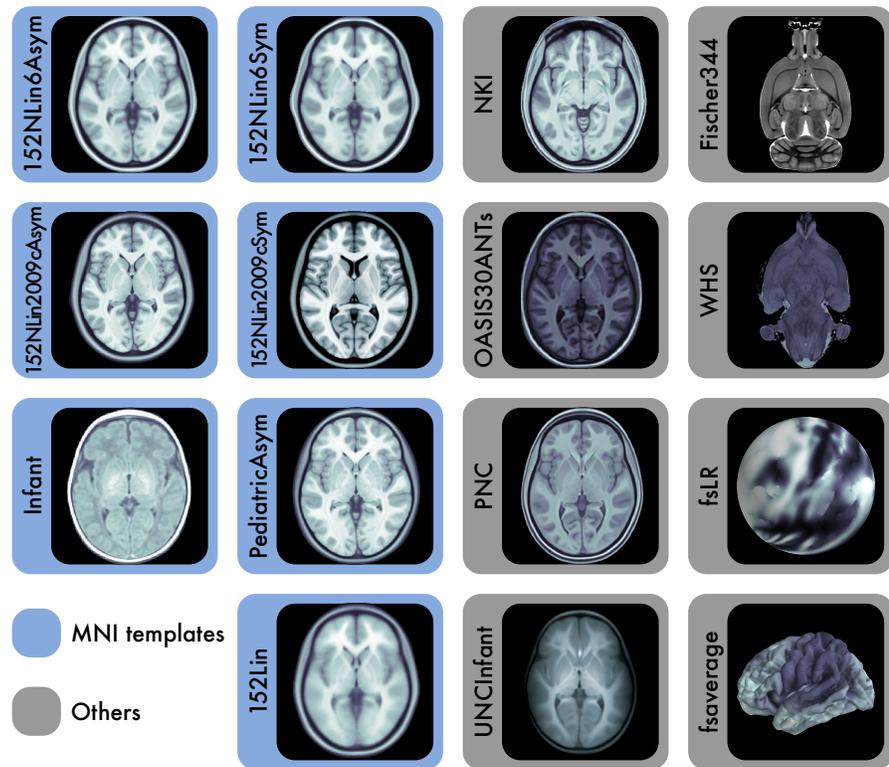
57 Because they are integral to analytic workflows, templates and atlases are frequently distributed  
58 as part of neuroimaging software libraries. For the most part, the developers of these libraries  
59 have substantial commitments apart from template aggregation and curation; thus, most libraries  
60 are practically limited in the subset of templates and atlases that they include. As an unfortunate  
61 consequence of this distribution model, access to and reuse of templates and atlases has become  
62 tightly coupled to a user's choice of software library. As an alternative to this software-bound  
63 distribution model, some laboratories and institutions maintain repositories where templates and  
64 atlases can be downloaded. The Montreal Neurological Institute (MNI) has spearheaded this mode  
65 of distribution and offers a large portfolio of human and non-human templates and atlases (*Evans*

66 *et al., 1993; Mazziotta et al., 1995; Holmes et al., 1998; Collins et al., 1999; Mazziotta et al., 2001;*  
67 *Fonov et al., 2011*) accessible via a web site (*MNI, n.d.*). These templates and atlases have evolved  
68 iteratively (*Evans et al., 2012*), preserving spatial alignment to the “MNI Average Brain (305 MRI)  
69 Stereotaxic Registration Model” (“MNI305”; *Evans et al., 1993*). As a consequence, it is common  
70 to find references to them in the literature under the umbrella term of “*MNI space*”. Indeed, the  
71 default templates distributed with the popular FSL (*Jenkinson et al., 2012*) and SPM (*Friston et al.,*  
72 *2006*) software packages are generally referred to as being “*defined in MNI space,*” even though the  
73 specific templates differ.

74 The limitations of the software-bound distribution model underscore three separate problems  
75 that arise in common practice. First, software-default templates are not generalizable to many use  
76 cases. When the population targeted by a study substantially deviates from neurotypical human  
77 adults (e.g., infants, elderly, or nonhuman animals), using an inadequate reference such as the  
78 default *MNI space* offered by a software library can introduce so-called “template effects” that bias  
79 morphometric analyses and produce incorrect results (*Yoon et al., 2009*). There is not yet any stan-  
80 dard distance function that can objectively determine whether a template choice is phenotypically  
81 proximal to the study’s sample, and thus whether template effects will be relevant. For example,  
82 since most MNI templates are created with a sample of adults of European ancestry, a study in-  
83 volving East Asian adults might require a non-default template. Because of the relative scarcity of  
84 nonhuman imaging resources, exposure to template effects is even more pressing in the nonhu-  
85 man context: e.g., is it appropriate to use a mouse template for the spatial standardization of rat  
86 images? Not only are nonhuman templates and atlases scarce, accommodation of such resources  
87 in popular software tools is generally limited. For instance AFNI (*Cox and Hyde, 1997*) includes a  
88 rat template that can be applied in some contexts, while SPM provides functionality only through  
89 third-party add-ons (e.g., *Sawiak et al., 2009*). Second, deviating from software defaults places a  
90 knowledge burden on the user. Once the researcher has selected a reference standard space that  
91 is suitable for their study population, if their choice is not included by default with the software  
92 they plan to use, they must then locate and download the reference template or atlas and inte-  
93 grate it within their analytic pipeline. This kind of excursion from defaults is far from frictionless  
94 and will often require expertise in template spaces and pipeline informatics. The required exper-  
95 tise is greater still when a researcher is working with an under-represented population for which no  
96 suitable template currently exists. In this situation, researchers often develop and make available  
97 new templates and atlases based on their own data samples, afterward distributing the new data  
98 assets using institutional websites or data storage systems such as FigShare (RRID:SCR\_004328) or  
99 Dryad (RRID:SCR\_005910). The lack of a centralized index for such templates propagates a share of  
100 the knowledge burden to researchers who stand to benefit from reusing them. Users must instead  
101 be aware not only of the prior existence of a template, but also where to locate it and the methods  
102 required to access it. Finally, as illustrated by the case of *MNI space*, it is not always clear what tem-  
103 plate a study is using. Since the templates most often used in the literature are software library  
104 defaults, reporting of spatial standardization is generally implicit (e.g., *Carp, 2012b*, for functional  
105 MRI studies). In addition, template and atlas curators do not generally mint universally unique  
106 identifiers (such as the Research Resource Identifier, RRID; *Bandrowski and Martone, 2016*) to pre-  
107 cisely report spatial standardization and analysis. Therefore, deviating from software defaults has  
108 some potential to endanger reproducibility of studies due to template/atlas accessibility (and conti-  
109 nuity thereof through time) and the risk of misreporting. Additional concerns regarding the repro-  
110 ducibility of spatial standardization in research include unlicensed distribution and provenance  
111 tracking. Errors in template and atlas resources are not common, but have been reported (e.g.,  
112 *Rohlfing, 2013; Halchenko, 2013*). Using version control for templates and atlases has traditionally  
113 been considered too onerous and requires an expertise that may exceed the resources of research  
114 teams.

115 Overall, current practices in management and stewardship of group-standardized data (tem-  
116 plates, atlases, and associated resources) do not follow “Findability, Accessibility, Interoperability,

**Figure 1. Representative views of 15 templates currently available in the *TemplateFlow Archive*.** The 7 templates highlighted in blue are constituents of the Montreal Neurological Institute (MNI) portfolio. **WHS** (Waxholm space) and **Fischer344** correspond to rat templates. **fsaverage** and **fsLR** are surface templates; the remaining templates are volumetric. Each template is distributed with atlas labels, segmentations, and metadata files. The 15 templates displayed here are only a small fraction of those created as stereotaxic references for the neuroimaging community.



117 and Reusability (FAIR) Guiding Principles” (*Wilkinson et al., 2016*), making it difficult for researchers  
 118 to locate and use these data assets and thereby reducing their long-term value. To address the  
 119 need for a centralized resource for the archiving and redistribution of templates and atlases that  
 120 allows programmatic access to human and nonhuman resources, we have developed *TemplateFlow*.  
 121 This resource implements the FAIR Guiding Principles, effectively decoupling standardized  
 122 spatial data from software libraries while maximizing the flexibility of emerging processing and  
 123 analysis workflows (e.g. *Esteban et al., 2017, 2019*). *TemplateFlow* comprises a cloud-based repos-  
 124 itory of human and nonhuman imaging templates —the “*TemplateFlow Archive*”, Figure 1— paired  
 125 with a Python-based library —the “*TemplateFlow Client*”— for programmatically accessing template  
 126 resources. The resource is complemented with a “*TemplateFlow Manager*” tool to upload new or  
 127 update existing resources. When adding a new template, the *Manager* initiates a peer-reviewed  
 128 contribution pipeline where experts are invited to curate and vet new proposals. These software  
 129 components, as well as all template resources, are version-controlled. Therefore, not only does  
 130 *TemplateFlow* enable access to templates “off-the-shelf” by humans and machines, it also permits  
 131 researchers to share their resources with the community. To implement several of the FAIR Prin-  
 132 ciples, the *TemplateFlow Archive* features a tree-directory structure, metadata files, and data files  
 133 following an organization inspired by the Brain Imaging Data Structure (BIDS; *Gorgolewski et al.,*  
 134 *2016*). BIDS is a widespread standard that balances the needs for human- and machine-readability.  
 135 The online documentation hub and the resource browser located at [TemplateFlow.org](https://TemplateFlow.org) provide fur-  
 136 ther details for users.

**Table 1. Digital templates included in *TemplateFlow*.** *TemplateFlow* is designed to maximise the discoverability and accessibility of new templates, minimise redundancies in template creation, and promote standardisation of processing workflows. To enhance visibility of existing templates, *TemplateFlow* includes a web-based browser indexing all files in the *TemplateFlow Archive* ([templateflow.org/browse/](http://templateflow.org/browse/)).

Template ID	Description
MNI152Lin	Neurotypical adult human template created as the average from a linear mapping of 152 subjects from the MNI cohort of the ICBM registered to the earlier MNI305 template ( <i>Mazziotta et al., 1995, 2001</i> ).
MNI152NLin6Asym	FSL's version of the MNI152 neurotypical adult human template created using iterative nonlinear registration and averaging ( <i>Evans et al., 2012</i> ).
MNI152NLin6Sym	Symmetric version of <a href="#">MNI152NLin6Asym</a> ( <i>Evans et al., 2012</i> ).
MNI152NLin2009cAsym	Update of the MNI152 neurotypical adult template with nonlinear registration. The mapping and averaging proceeded over 40 iterations beginning from the earlier MNI152 template ( <i>Fonov et al., 2011; Collins et al., 1999</i> ).
MNI152NLin2009cSym	Symmetric version of <a href="#">MNI152NLin2009cAsym</a> ( <i>Fonov et al., 2011; Collins et al., 1999</i> ).
MNIInfant	Series of human infant templates created from 11 cohorts of infants and young children. Each cohort spans a different age range between 0 and 60 months ( <i>Fonov et al., 2011</i> ).
MNIPediatricAsym	Series of human pediatric templates created from 6 partially overlapping cohorts of children and young adults. Each cohort spans a different age range between 4.5 and 18.5 years ( <i>Fonov et al., 2011; Collins et al., 1999</i> ).
NKI	Template created for the NKI-Rockland sample using ANTs diffeomorphic registration and averaging ( <i>Avants et al., 2011; Nooner et al., 2012</i> ).
OASIS30ANTs	Template created using ANTs diffeomorphic registration and averaging for the Open Access Series of Imaging Studies (OASIS) ( <i>Avants et al., 2011; Marcus et al., 2007</i> ).
PNC	Pediatric and young adult template created using ANTs diffeomorphic registration and averaging for the Philadelphia Neurodevelopmental Cohort ( <i>Satterthwaite et al., 2016</i> ).
UNCInfant	Series of human infant templates created from a 95-subject longitudinal sample comprising three scans: as neonates, as one-year-olds, and as two-year-olds ( <i>Shi et al., 2011</i> ).
WHS	Waxholm space template created as an atlas of the Sprague-Dawley rat brain ( <i>Kjonigsen et al., 2015; Osen et al., 2019; Papp et al., 2014, 2015</i> ).
Fischer344	Rat template created as the average of 41 four-month-old animals from the Fischer 344 strain ( <i>Goerzen et al., 2020</i> ).
fsLR	Surface-based Freesurfer template created for the Human Connectome Project (HCP) ( <i>Van Essen et al., 2012</i> ).
fsaverage	Surface-based average Freesurfer template ( <i>Fischl et al., 1999</i> ).

## 137 Results

### 138 Management and stewardship following the FAIR Guiding Principles

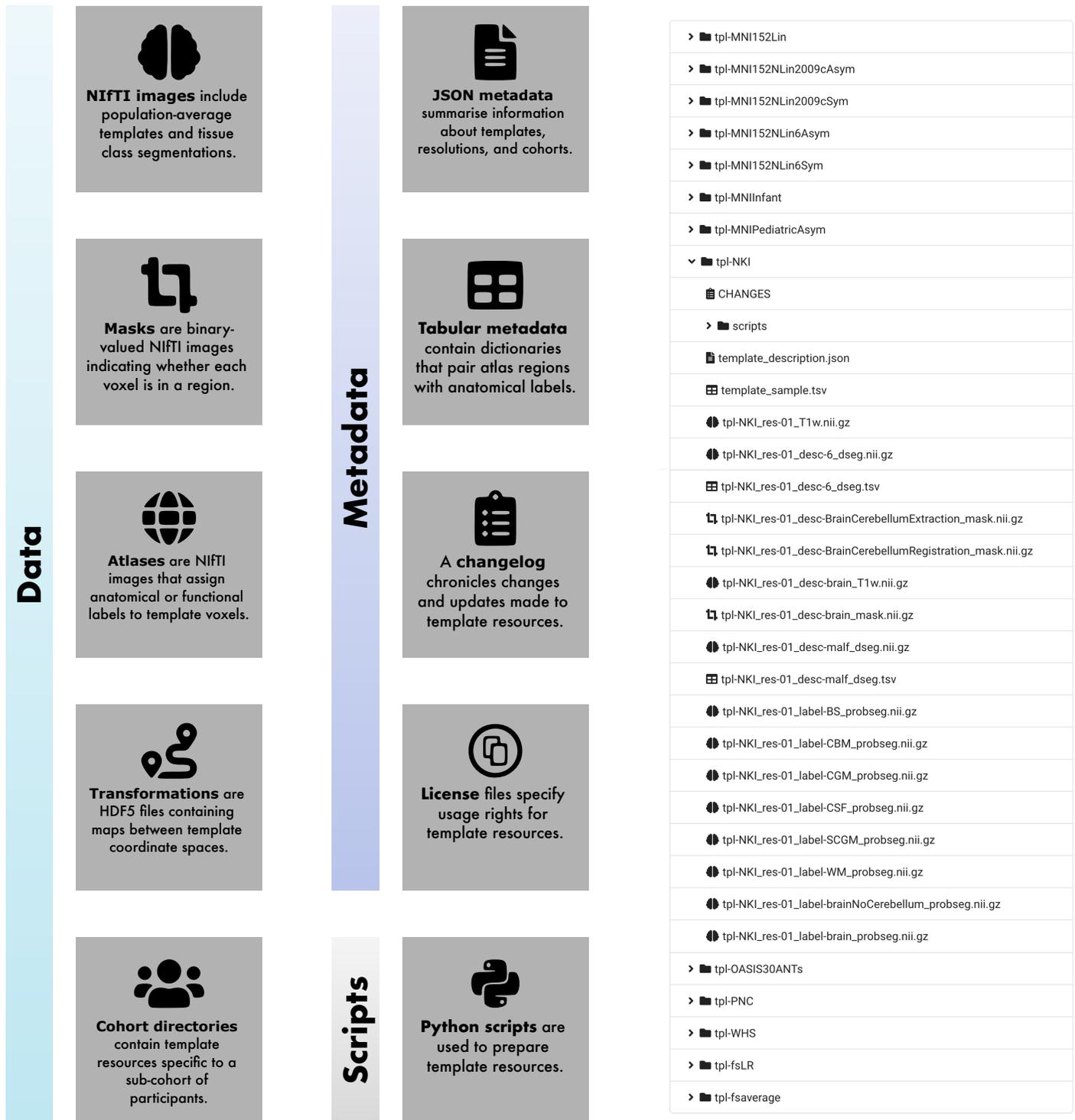
139 The specific and measurable FAIR principles of *Wilkinson et al. (2016)* are reproduced in Sup-  
140 plementary Box S1. Principles are indexed by their category and numbering, e.g., principle I3 –  
141 “(meta)data include qualified references to other (meta)data” belongs in the “to be Interoperable” cat-  
142 egory. We describe how *TemplateFlow* implements each of these specific principles in the follow-  
143 ing. Every template and all associated metadata, atlases, etc. are assigned a unique and persistent  
144 identifier (principle F1). BIDS prescribes a file naming scheme comprising a series of key-value pairs  
145 (called “entities”) that are ordered hierarchically. Following BIDS’ patterns, the template identifier  
146 is an alphanumeric label that is unique across the *Archive*, signified with the key `tpl-` (e.g., `tpl-`  
147 `MNI152Lin`). *TemplateFlow* therefore adapts BIDS to the specific domain of templates and atlases,  
148 affording the tool with a robust implementation of the principles I1-3 (interoperability). Table 1 enu-  
149 merates several templates currently distributed with the *Archive*, and their corresponding unique  
150 identifiers. The unique identifier resolves the issue of inaccurate reporting, as it unambiguously  
151 designates one specific template. In addition, because the repository is versioned, researchers can  
152 easily retrieve and report the exact version of the template or atlas that was used in their study.  
153 Suppl. Table S2 summarizes the available entities and shows a segment of the file organization  
154 of the *Archive*. For each template, the *TemplateFlow* database includes reference volumetric tem-  
155 plate images (e.g., one T1-weighted and one T2-weighted average map; all must be in register), a  
156 set of atlas labels and voxelwise annotations defined with reference to the template image, and  
157 additional files containing the template and atlas metadata. Correspondingly, *TemplateFlow* allows  
158 surface-based resources such as average features, geometry files, annotations, or metadata.

159 Template resources are described with rich metadata (principles F2 and R1), ensuring that the  
160 data usage license is clear and accessible (R1.1), data and metadata are associated with detailed  
161 provenance (R1.2), and data and metadata follow a domain-relevant structure transferred from  
162 the neuroimaging community standards of BIDS (R1.3). Figure 2 summarizes the data types and  
163 metadata that can be stored in the *Archive*. Figure 3 provides an overview of the *Archive*’s metadata  
164 specification, showing that metadata clearly and explicitly include the identifier of the data they de-  
165 scribe (F3). Data and metadata are retrievable using several open, free, standard communications  
166 protocols without need for authentication (A1) by using *DataLad* (*Halchenko et al., 2021*). Cloud  
167 storage for the *Archive* is supported by the Open Science Framework (osf.io) and Amazon’s Simple  
168 Storage Service (S3). Version control, replication, and synchronisation of template resources across  
169 filesystems is managed with *DataLad*. Leveraging *DataLad*, metadata are stored on GitHub, ensur-  
170 ing accessibility to metadata even when corresponding data are no longer available (A2). *DataLad*  
171 is based on *Git* and *Git-Annex*, which index all data and metadata (F4). Although *DataLad* also pro-  
172 vides searching tools that are applicable to *TemplateFlow*’s resources, the *TemplateFlow* framework  
173 provides a client tool that facilitates searching and querying.

### 174 An indexed resource, searchable with a *TemplateFlow* “Client”

175 *TemplateFlow*’s Python client provides human users and software tools with reliable and program-  
176 matic access to the archive. The client can be integrated seamlessly into image processing work-  
177 flows to handle requests for template resources on the fly. It features an intuitive application pro-  
178 gramming interface (API) that can query the *TemplateFlow Archive* for specific files (Figure 5). The  
179 BIDS-inspired organization enables easy integration of tools and infrastructure designed for BIDS  
180 (e.g., the Python client uses *PyBIDS* (*Yarkoni et al., 2019*) to implement queries like those listed in  
181 Suppl. Table S2). To query *TemplateFlow*, a user can submit a list of arguments corresponding to  
182 the BIDS-like key-value pairs in each entity’s file name (e.g., `atlas="Schaefer2018"` to return files  
183 containing voxelwise annotations by *Schaefer et al. (2018)*).

184 To integrate template resources into neuroimaging workflows, traditional approaches required  
185 deploying an oftentimes voluminous tree of prepackaged data to the filesystem. By contrast, the  
186 *TemplateFlow* client implements lazy loading, which permits the base installation to be extremely



**Figure 2. The *TemplateFlow Archive* contains template resources.** Left, common file formats included in the *TemplateFlow Archive*. Right, view of the *TemplateFlow Archive*'s browser, accessible at [TemplateFlow.org](https://TemplateFlow.org), with a single template resource directory expanded. Template data are archived using a BIDS-like directory structure, with top-level directories for each template. Each directory contains image files, annotations, and metadata for that template. Following BIDS specifications, volumetric data are stored in NIfTI-1 format. Further surface-based data types are supported with GIFTI (surfaces) and CIFTI-1 (mixed volumetric-and-surface data).

```

{
  "Authors": [
    "Fonov V",
    "Evans AC",
    "Botteron K",
    "Almli CR",
    "McKinstry RC",
    "Collins DL"
  ],
  "Curators": [
    "Esteban O"
  ],
  "Identifier": "MNIPediatricAsym",
  "License": "MIT-derived. See LICENSE file",
  "Name": "MNI's unbiased standard MRI template for
pediatric data from the 4.5 to 18.5y age
range",
  "RRID": "SCR_008796",
  "ReferencesAndLinks": [
    "https://doi.org/10.1016/j.neuroimage.2010.07.033",
    "https://doi.org/10.1016/S1053-8119(09)70884-5",
    "http://nist.mni.mcgill.ca/?p=974",
    "https://doi.org/10.1007/3-540-48714-X_16"
  ],
  "TemplateFlowVersion": "1.0.0",
  "cohort": {
    "1": {
      "age": [
        4.5,
        18.5
      ],
      "name": "whole age range",
      "units": "yr"
    },
    "2": {
      "age": [
        4.5,
        8.5
      ],
      "name": "prepuberty",
      "units": "yr"
    },
    . . .
  },
  "res": {
    "1": {
      "origin": [
        -98.0,
        -134.0,
        -72.0
      ],
      "shape": [
        197,
        233,
        189
      ],
      "zooms": [
        1.0,
        1.0,
        1.0
      ]
    },
    . . .
  }
}

```

## General metadata

Field	Type	Description
Authors	Array (String)	Names of authors who created the template.
Curators	Array	Names of TemplateFlow curators who contributed or manage the dataset.
Identifier	String	Unique human-readable template identifier within TemplateFlow.
License	String	License under which template resources are available.
Name	String	Full descriptive name of the template.
RRID	String	Research Resource Identifier for the TemplateFlow dataset.
ReferencesAndLinks	Array	Publications to reference when using the template, and salient links for template information.
TemplateFlowVersion	String	Version of TemplateFlow under which the dataset was distributed.

## Cohort metadata

Field	Type	Description
cohort	Object	Top-level field containing all cohort metadata objects.
1, 2, ...	Object	Cohort identifiers. Each has a subdirectory in the template data directory, and each has a metadata object nested in the cohort field.
age	Array (Number)	2-tuple array indicating the lower and upper bounds for participant age in the cohort, if the cohorts are stratified by age.
name	String	Full descriptive name of the cohort..
units	String	Units for cohort age bounds.

## Resolution metadata

Field	Type	Description
res	Object	Top-level field containing all resolution metadata objects.
1, 2, ...	Object	Resolution identifiers. Each has a metadata object nested in the res field. The metadata for each resolution apply to all images whose name includes res- <i>&lt;identifier&gt;</i> . The identifier itself does not necessarily correspond to the voxel size.
origin	Array	(x, y, z) spatial location of the voxel origin relative to the physical origin in mm.
shape	Array	(x, y, z) shape of the image in voxels.
zooms	Array	(x, y, z) size of each voxel in mm.

**Figure 3. Overview of the metadata specification of the *TemplateFlow Archive*.** *TemplateFlow's* metadata are formatted as JavaScript Object Notation (JSON) files located within each template set. An example `template_description.json` metadata file is displayed at the left for `MNIPediatricAsym`. In addition to general template metadata, datasets can contain cohort-level and resolution-level metadata, which are nested within the main metadata dictionary and apply only to subsets of images in the dataset.

187 lightweight. Instead of distributing neuroimaging data with the installation, *TemplateFlow* allows  
188 the user to dynamically pull from the cloud-based storage only those resources they need, as they  
189 need them. After a resource has been requested once, it remains cached in the filesystem for  
190 future utilization.

191 We demonstrate benefits of centralizing templates in general, and the validity of the *TemplateFlow*  
192 *Flow* framework in particular, via its integration into *fMRIPrep* (Esteban et al., 2019), a functional  
193 MRI preprocessing tool. This integration provides *fMRIPrep* users with flexibility to spatially nor-  
194 malize their data to any template available in the *Archive* (see Box 1). This integration has also  
195 enabled the development of *fMRIPrep* adaptations, for instance to pediatric populations or rodent  
196 imaging (MacNicol et al., 2021), using suitable templates from the archive. The uniform interface  
197 provided by the BIDS-like directory organisation and metadata enables straightforward integra-  
198 tion of new templates into workflows equipped to use *TemplateFlow* templates. Further examples  
199 of tools leveraging *TemplateFlow* include *MRIQC* (Esteban et al., 2017) for quality control of MRI;  
200 *PyNets* (Pisner and Hammonds, 2020), a package for ensemble learning of functional and struc-  
201 tural connectomes; *ASLPrep* (Adebimpe et al., 2021), an ASL pre-processing pipeline that makes  
202 use of *TemplateFlow* through *sMRIPrep* —the spin-off structural pipeline from *fMRIPrep*; and *Net-*  
203 *PlotBrain* (Thompson and Fanton, 2021), which uses *TemplateFlow* to display spatially standardized  
204 brain network data.

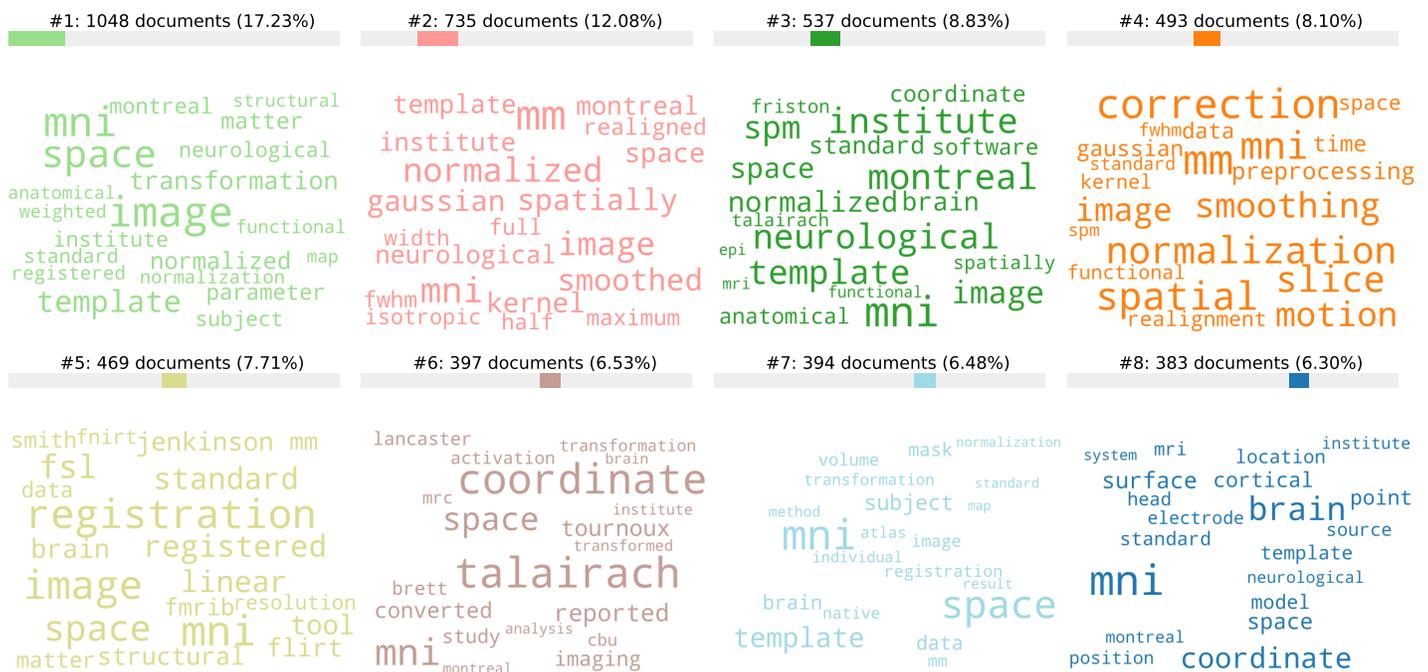
### 205 **A framework for researchers who generate and share spatially-standardized data**

206 A centralized repository for neuroimaging templates should also address the needs of template  
207 creators, enabling peer-reviewed integration of new templates with minimal informatic overhead.  
208 Inspired by the Conda-forge community repository and the Journal of Open Source Software, the  
209 GitHub-based “*templateflow*” organisation is a site for dialogue between members of the neu-  
210 roimaging community and *TemplateFlow Archive* curators. GitHub issues offer any community mem-  
211 ber the ability to share their needs with developers and *Archive* curators, for instance by identify-  
212 ing templates or workflow features for potential inclusion in the project. “Pull requests” provide  
213 a means for members of the community to directly contribute code or template resources to the  
214 *TemplateFlow Archive*.

215 This peer-reviewed contribution process is facilitated through the Python-based *TemplateFlow*  
216 *Manager*. The *TemplateFlow Manager* automates the work of synchronizing data from a local di-  
217 rectory to cloud storage in OSF. Furthermore, it creates a GitHub repository containing git-annex  
218 pointers that enable *DataLad* to download template data from cloud storage to any machine with a  
219 copy of the repository. Finally, it opens a new pull request to propose adding the newly contributed  
220 template repository into the main *TemplateFlow Archive* (Figure 6). Synchronization of spatial data  
221 assets to the *TemplateFlow Archive* affords data producers an immediate way to distribute their  
222 data according to FAIR principles and thereby increase its reach.

### 223 **Unambiguous and precise reporting of spatially standardized processing and analysis.**

224 To explore the coupling between software libraries and standard spaces, we conducted a topic  
225 modeling analysis (Blei et al., 2003) of *MNI* space in the neuroimaging literature. We identified a  
226 coupling between the reporting of spatial standardization and software libraries across 6,048 arti-  
227 cles containing the term *MNI* and published in two leading domain-specific journals (*NeuroImage*  
228 and *NeuroImage: Clinical*). To demonstrate the heterogeneity in the reported standardization to  
229 *MNI space*, we sorted topics according to their dominance in articles (i.e., the topic with the high-  
230 est model score in *MNI*-related sentences; Figure 4). Out of 15 topics we modeled, two of the  
231 most dominant topics contained software tool names as well as the names of related scientists.  
232 As shown in Figure 4, around 500 articles (each term) contained either “SPM” (9% of the docu-  
233 ments) or “FSL” (8%). Interestingly, the two words do not ever appear together, suggesting that  
234 researchers stick with one or another in their analyses. Additional topics that seemingly relate to  
235 the provenance of templates and atlases —beyond the ubiquitous “Montreal”, “Neurological” and



**Figure 4. The FSL and SPM software tools associate with dominant topics of sentences including the term “MNI” across the literature.** We performed topic modeling with latent Dirichlet allocation (LDA; *Blei et al., 2003*) on text sentences extracted from 6,048 articles that contained the word “MNI”. For each topic identified, the 20 words with the highest loadings on that topic are displayed in a word cloud with larger font size indicating higher loading of the word on the corresponding topic. Word clouds are sorted by descending topic dominance. Ranking and relative dominance are shown above each topic’s cloud. Only the top 8 dominant topics are shown here (full model is reported in Suppl. Figure S1). Two top-dominant topics —#3 and #5— are associated with SPM and FSL respectively.

236 “Institute” for MNI— are those that ranked #4, #13, #14, which include “SPM” and other terms such  
 237 as “McGill”, “Wellcome”, “UCL”, or the “parametric” in SPM (see Suppl. Figure S1). The remainder  
 238 of topics appears to relate to miscellaneous aspects of spatial standardization, such as “anatomical”,  
 239 “smoothness”, “map/mapping”, “standard”, “normalization”, “(re)align/alignment”, etc., with  
 240 no apparent relationship to the actual origin of the resource. These interpretations suggest that  
 241 “MNI space” can refer to any of a family of templates and is not a unique identifier. As a matter of  
 242 fact, studies carried out with SPM96 (*Friston et al., 2006*) and earlier versions report their results  
 243 in *MNI space* with reference to the single-subject Colin 27 average template (*Holmes et al., 1998*).  
 244 However, beginning with SPM99, SPM updated its definition of *MNI space* to the template that MNI  
 245 released in 2001: an average of 152 subjects from the ICBM database, aligned by means of linear  
 246 registration. In SPM12 (the latest release at the time of writing), the meaning of *MNI template*  
 247 varies by submodule: different modules alternately use the Linear MNI152 template (*Mazziotta*  
 248 *et al., 1995*) and a new, nonlinear revision from 2009 (*Fonov et al., 2011*). By contrast, the *MNI*  
 249 *template* bundled with the FSL toolbox was developed by Dr. A. Janke in collaboration with MNI  
 250 researchers (*Evans et al., 2012*). Although it was generated under the guidance of and using the  
 251 techniques of the 2006 release of nonlinear MNI templates, this template is not in fact part of the  
 252 official portfolio distributed by MNI. Nonetheless, our results suggest that the MNI templates bundled  
 253 with SPM and FSL have historically gained broader currency as a result of the widespread use  
 254 of these software libraries.

## 255 Discussion

256 The use of templates and atlases is ubiquitous in neuroimaging, and the emerging challenges regarding  
 257 template use accordingly merit immediate attention. In an early perspective, Van Essen  
 258 identified a set of desiderata for brain templates (*Van Essen, 2002*). Above and beyond anatomical

259 fidelity, he called for connecting templates in an aggregation of databases with “powerful and flex-  
260 ible options for searching, selecting, and visualizing data”. Finally, he stressed the importance of  
261 resource accessibility. *TemplateFlow* provides a framework that satisfies all of the aforementioned  
262 desiderata while following the FAIR principles (*Wilkinson et al., 2016*). We elaborate that most of  
263 the issues concerning the reliability of neuroimaging research relating to standardized spatial ref-  
264 erences stem from the lack of a centralized repository designed to meet FAIR principles. We show  
265 how we effectively implement such principles with the adoption of a BIDS-like structure for the  
266 data and metadata in the resource, and with *DataLad* to support the core of the data management  
267 system. We complete the implementation with an easy-to-use client tool.

268 When researchers develop a new brain template or atlas for public dissemination, there exists  
269 no standard channel or format for distributing their work. With no central repository or uniform  
270 organizational scheme, template creators are often tasked with the responsibilities of maintain-  
271 ing template resources and managing access on an ad hoc basis and sustaining them over time  
272 with limited to no support. While the quality of peer reviewed template resources is assessed  
273 once prior to publication, reviewers often focus on perceived academic merit to the exclusion of  
274 FAIR principles. This can lead to poor resource adoption and low community value even for high-  
275 quality resources. Informal vetting of resources is prone to more clerical errors such as missing  
276 or corrupted files and/or metadata, or unlicensed distribution, which nonetheless may make the  
277 resource unusable. Conversely, users are confronted with a surfeit of available templates and  
278 atlases, many with unclear provenance, absent licensing terms, and the attendant challenges of  
279 accessing them and integrating them into workflows.

280 Without a uniform distribution format, integrating a template into software requires a custom  
281 solution for every new template, increasing the burden on developers. Consistently with the pre-  
282 vious investigation by *Carp (2012b)* in the domain of functional MRI, our text mining exploration  
283 illustrates a strong coupling between software library and the templates and atlases of choice. In-  
284 deed, *Carp (2012b)* analysed 241 functional MRI studies of which 90.9% reported normalizing brain  
285 images to a common template. Of those, 79.0% indicated the target space used for spatial normal-  
286 ization. Few studies reported critical parameters, and only 50 specified the template image: 26.0%  
287 used “the MNI152 template”, and 26.0% the “SPM library’s echo-planar imaging template”. Unfortu-  
288 nately, template selection is seen as a default parameter of the software library, which lends itself  
289 to assuming that the target normalization space is implicitly reported by identifying the software  
290 tools of choice. *Bowring et al. (2019)* wrote comparable pipelines in three software suites (AFNI,  
291 FSL and SPM) in order to identify challenges to reproducing published studies with openly shared  
292 raw data. When discussing differences among pipelines, they noted that, “while all packages are  
293 purportedly using the same MNI atlas space, an appreciable amount of activation detected by AFNI  
294 and FSL fell outside of SPM’s analysis mask”.

295 This coupling seems to also limit the utilization of templates other than those defaulted by the  
296 software. Custom templates (i.e., those not included as default option for the software tool) range  
297 from population-specific templates to ad hoc templates created by averaging images of the study  
298 at hand. In some settings, the use of default templates risks introducing “template effects” that  
299 confound the interpretation of results (such as those introduced in pediatric imaging studies by  
300 using an adult template, *Yoon et al., 2009*). As the target population moves away from that with  
301 which a default template was created, “template effects” become more concerning and custom  
302 templates more necessary. The problem is exacerbated in the case of nonhuman imaging, as the  
303 scarcity (or absence) of specific templates available within software packages hinders already chal-  
304 lenging translational endeavors. Further, the consistency across templates and atlases is report-  
305 edly low (*Bohland et al., 2009*), and although there has not been any programmatic comparison  
306 to understand the extent to which this inconsistency alters the spatial interpretation of results, it  
307 is reasonable that templates and atlases introduce a decision point and therefore are sources of  
308 some analytical variability.

309 One ostensible caveat regarding centralized and FAIR-principled knowledge repositories such

310 as *TemplateFlow* is that, by increasing the findability and reuse potential of data resources, they  
311 also open the door to increased methodological flexibility. *Carp* empirically investigated the con-  
312 sequences of such methodological flexibility in neuroimaging, demonstrating that decision points  
313 in workflows can lead to substantial variability in analysis outcomes. In a contemporaneous paper,  
314 *Carp (2012b)* contextualized these findings vis-à-vis the inflated risk of false positives, underscoring  
315 that analytical variability degrades the reproducibility of studies only *in combination with* (intended  
316 or unintended) *selective reporting* of methods and results. Selective reporting, in this particular ap-  
317 plication, would mean that a researcher explores the results with reference to several templates  
318 or atlases and reports only those that confirm the research's hypotheses. *TemplateFlow's* standard-  
319 ization preempts the problem of *unintended* selective reporting: the provenance of all resources is  
320 tracked, all resources are accessible, and comprehensive metadata are generated.<sup>13</sup> More recently,  
321 *Botvinik-Nezer et al. (2020)* advocated for another solution to the problem of analytical variability:  
322 “multiverse” analyses, wherein many combinations of methodological choices are all thoroughly re-  
323 ported and cross-compared when presenting results. Applied to the particular choice of template  
324 and atlas combinations, it would thus be desirable to report neuroimaging results with reference to  
325 several standard spaces and determine whether the interpretations hold across those references  
326 and atlases. *TemplateFlow's* interoperability empowers users to incorporate this type of analysis  
327 into their research by easily making template or atlas substitutions for cross-comparison. For in-  
328 stance, *Box 1* shows how *TemplateFlow* works with *fMRIPrep* to automate preprocessing of outputs  
329 in multiple standard spaces. This facilitates assessment of the robustness of a result with respect  
330 to the template or atlas of choice in accordance with the multiverse approach.

### 331 Limitations

332 *TemplateFlow* affords researchers substantial analytical flexibility in the choice of standard spaces  
333 of reference. Such flexibility helps researchers minimize “template effects” —by easily inserting the  
334 most adequate template— but also opens opportunities for incomplete reporting of experiments.  
335 Using *DataLad* or the *TemplateFlow Client*, researchers have at their disposal the necessary tooling  
336 for precise reporting: unique identifiers, provenance tracking, version tags, and comprehensive  
337 metadata. Therefore, the effectiveness of *TemplateFlow* to mitigate selective reporting is bounded  
338 by the user's discretion. Similarly, the resource is limited at the time of writing to MRI templates,  
339 but it readily supports such other modalities as nuclear imaging maps (e.g., PET/SPECT), and would  
340 support others with minimal adaptations (e.g., protein expression maps, or 3D reconstructions  
341 from histology).

342 As a research resource, the scope of this manuscript is limited to describing the framework and  
343 infrastructure of *TemplateFlow*, highlighting how neuroscientists can leverage this new data archive  
344 and the tooling around it. Therefore, some fundamental issues related to this work must be left for  
345 future investigation: (i) the overarching problems of cross-template and cross-atlas consistency;  
346 (ii) comparative evaluation of methodological alternatives for producing new templates, atlases  
347 and related data; (iii) providing neuroimagers with more objective means to determine the most  
348 appropriate template and atlas choices that apply to their research, as well as better understanding  
349 “template effects”; (iv) the adequacy of original (MRI, nuclear imaging, etc.) and derived (regularly  
350 gridded images, surfaces, etc.) modalities for a specific research application; or (v) the study of the  
351 validity and reliability of inter-template registration, as well as the evaluation of such a component  
352 of the *TemplateFlow* framework.

353 In more practical terms, *TemplateFlow* is limited to the redistribution of resources under per-  
354 missive licenses, without any access restrictions.

### 355 Conclusion

356 We introduce an open framework for the archiving, maintenance and sharing of neuroimaging  
357 templates and atlases called *TemplateFlow* that is implemented under FAIR data sharing principles.  
358 We describe the current need for this resource in the domain of neuroimaging, and further discuss  
359 the implications of the increased analytical flexibility this tool affords. These two facets of repro-

360 ducibility —availability (under FAIR guiding principles) of prior knowledge required by the research  
361 workflow, and the analytical flexibility such availability affords— are ubiquitous concerns across  
362 disciplines. *TemplateFlow's* approach to addressing both establishes a pattern broadly transfer-  
363 able beyond neuroimaging. We envision *TemplateFlow* as a core research tool undergirding multi-  
364 verse analyses —assessing whether neuroimaging results are robust across population-wide spa-  
365 tial references— as well as a stepping stone towards the quest of mapping anatomy and function  
366 across species.

### 367 **Acknowledgments**

368 The development of this resource was supported by the Laura and John Arnold Foundation (RAP  
369 and KJG), the NIBIB (R01EB020740, SSG; 1P41EB019936-01A1 SSG, YOH), NIMH (RF1MH121867 RAP,  
370 OE; R24MH114705 and R24MH117179, RAP; 1RF1MH121885 SSG), NINDS (U01NS103780, RAP),  
371 and NSF (CRCNS 1912266, YOH). RL is funded by the Wellcome Trust (209139/Z/17/Z). EM was  
372 supported by the UK Medical Research Council (MR/N013700/1) and King's College London. OE ac-  
373 knowledges financial support from the SNSF Ambizione project “Uncovering the interplay of struc-  
374 ture, function, and dynamics of brain connectivity using MRI” (grant number PZ00P2\_185872).

### 375 **Ethical compliance**

376 We complied with all relevant ethical regulations. This resource reused publicly available data de-  
377 rived from studies acquired at many different institutions. Protocols for all of the original studies  
378 were approved by the corresponding ethical boards.

### 379 **Code & data availability statement**

380 All the software components discussed in this paper are available under the Apache 2.0 license,  
381 accessible as repositories of <https://github.com/templateflow>. All templates and associated data  
382 are available under corresponding open licenses and accessible as described in the manuscript.

### 383 **Author contributions**

384 Conceptualization: RC, CJM, KJG, RAP, OE Data curation: RC, EM, OE Topics analysis: RC, RL, OE  
385 Funding acquisition: KJG, RAP, OE Methodology: RC, OE Project administration: RAP, OE Resources:  
386 YOH, SSG, KJG, RAP, OE Software & Documentation: RC, WHT, MG, EM, CJM, YOH, OE Supervision:  
387 RAP, OE Validation: RC, WHT, EM, OE Visualization: RC, RL. Writing – original draft: RC & OE Writing  
388 – review & editing: RC, RL, WHT, MG, EM, CJM, YOH, SSG, KJG, RAP, OE.

### 389 **References**

- 390 **Adebimpe A**, Bertolero M, Dolui S, Cieslak M, Murtha K, Baller EB, Boeve B, Boxer A, Butler ER, Cook P, Colcombe  
391 S, Covitz S, Davatzikos C, Davila DG, Elliott MA, Flounders MW, Franco AR, Gur RE, Gur RC, Jaber B, et al.  
392 ASLPrep: A Generalizable Platform for Processing of Arterial Spin Labeled MRI and Quantification of Regional  
393 Brain Perfusion. *bioRxiv*. 2021 May; p. 2021.05.20.444998. doi:10.1101/2021.05.20.444998.
- 394 **Avants BB**, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-  
395 correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008;  
396 12(1):26–41. doi:10.1016/j.media.2007.06.004.
- 397 **Avants BB**, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs  
398 similarity metric performance in brain image registration. *NeuroImage*. 2011 Feb; 54(3):2033–44.  
399 doi:10.1016/j.neuroimage.2010.09.025.
- 400 **Bandrowski AE**, Martone ME. RRIIDs: A Simple Step toward Improving Reproducibility through Rigor and Trans-  
401 parency of Experimental Methods. *Neuron*. 2016 May; 90(3):434–436. doi:10.1016/j.neuron.2016.04.030.
- 402 **Blei DM**, Ng AY, Jordan MI. Latent Dirichlet Allocation. *J Mach Learn Res*. 2003; 3(Jan):993–1022. [https://jmlr.](https://jmlr.org/papers/v3/blei03a.html)  
403 [org/papers/v3/blei03a.html](https://jmlr.org/papers/v3/blei03a.html).

- 404 **Bohland JW**, Bokil H, Allen CB, Mitra PP. The Brain Atlas Concordance Problem: Quantitative Comparison of  
405 Anatomical Parcellations. *PLoS One*. 2009 Sep; 4(9):e7200. doi:10.1371/journal.pone.0007200.
- 406 **Botvinik-Nezer R**, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, Kirchler M, Iwanir R, Mumford  
407 JA, Adcock RA, Avesani P, Baczkowski BM, Bajracharya A, Bakst L, Ball S, Barilari M, Bault N, Beaton D, Beitner  
408 J, Benoit RG, et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*. 2020  
409 Jun; 582(7810):84–88. doi:10.1038/s41586-020-2314-9.
- 410 **Bowring A**, Maumet C, Nichols TE. Exploring the impact of analysis software on task fMRI results. *Human Brain*  
411 *Mapping*. 2019; 40(11):3362–3384. doi:https://doi.org/10.1002/hbm.24603.
- 412 **Brett M**, Johnsrude IS, Owen AM. The problem of functional localization in the human brain. *Nat Rev Neurosci*.  
413 2002; 3:243–249. doi:10.1038/nrn756.
- 414 **Brodmann K**. Brodmann's: Localisation in the Cerebral Cortex. Springer US; 2006. doi:10.1007/b138298.
- 415 **Buckner RL**, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, Snyder AZ. A unified approach for morphometric  
416 and functional data analysis in young, old, and demented adults using automated atlas-based head size nor-  
417 malization: reliability and validation against manual measurement of total intracranial volume. *NeuroImage*.  
418 2004 Oct; 23(2):724–738. doi:10.1016/j.neuroimage.2004.06.018.
- 419 **Calabrese E**, Badea A, Watson C, Johnson GA. A quantitative magnetic resonance histology atlas of postnatal  
420 rat brain development with regional estimates of growth and variability. *NeuroImage*. 2013 May; 71:196–206.  
421 doi:10.1016/j.neuroimage.2013.01.017.
- 422 **Carp J**. On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of fMRI Experiments.  
423 *Front Neurosci*. 2012; 6. doi:10.3389/fnins.2012.00149.
- 424 **Carp J**. The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage*. 2012 Oct;  
425 63(1):289–300. doi:10.1016/j.neuroimage.2012.07.004.
- 426 **Coalson TS**, Van Essen DC, Glasser MF. The impact of traditional neuroimaging methods on the spatial lo-  
427 calization of cortical areas. *Proceedings of the National Academy of Sciences*. 2018; 115(27):E6356–E6365.  
428 doi:10.1073/pnas.1801582115.
- 429 **Collins DL**, Zijdenbos AP, Baaré WFC, Evans AC. ANIMAL+INSECT: Improved Cortical Structure Segmentation.  
430 In: Kuba A, Šáamal M, Todd-Pokropek A, editors. *Information Processing in Medical Imaging Lecture Notes in*  
431 *Computer Science*, Berlin, Heidelberg: Springer; 1999. p. 210–223. doi:10.1007/3-540-48714-X\_16.
- 432 **Courchesne E**, Chisum HJ, Townsend J, Cowles A, Covington J, Egaas B, Harwood M, Hinds S, Press GA. Normal  
433 Brain Development and Aging: Quantitative Analysis at in Vivo MR Imaging in Healthy Volunteers. *Radiology*.  
434 2000 Sep; 216(3):672–682. doi:10.1148/radiology.216.3.r00au37672.
- 435 **Cox RW**, Hyde JS. Software tools for analysis and visualization of fMRI data. *NMR Biomed*. 1997; 10(4-5):171–  
436 178. doi:10.1002/(SICI)1099-1492(199706/08)10:4<171::AID-NBM453>3.0.CO;2-L.
- 437 **Dickie DA**, Job DE, Gonzalez DR, Shenkin SD, Wardlaw JM. Use of Brain MRI Atlases to Determine Bound-  
438 aries of Age-Related Pathology: The Importance of Statistical Method. *PLoS One*. 2015 May; 10(5):e0127939.  
439 doi:10.1371/journal.pone.0127939.
- 440 **Dickie DA**, Shenkin SD, Anblagan D, Lee J, Blesa Cabez M, Rodriguez D, Boardman JP, Waldman A, Job DE,  
441 Wardlaw JM. Whole Brain Magnetic Resonance Image Atlases: A Systematic Review of Existing Atlases and  
442 Caveats for Use in Population Imaging. *Front Neuroinform*. 2017; 11. doi:10.3389/fninf.2017.00001.
- 443 **Esteban O**, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: Advancing the Au-  
444 tomatic Prediction of Image Quality in MRI from Unseen Sites. *PLoS One*. 2017 Aug; 12(9):e0184661.  
445 doi:10.1371/journal.pone.0184661.
- 446 **Esteban O**, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M,  
447 Oya H, Ghosh SS, Wright J, Durnez J, Poldrack RA, Gorgolewski KJ. fMRIPrep: a robust preprocessing pipeline  
448 for functional MRI. *Nat Meth*. 2019 Jan; 16(1):111–116. doi:10.1038/s41592-018-0235-4.
- 449 **Evans AC**, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM. 3D statistical neuroanatomical models from 305  
450 MRI volumes. In: *IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, vol. 3  
451 San Francisco, CA, USA; 1993. p. 1813–1817. doi:10.1109/NSSMIC.1993.373602.

- 452 **Evans AC**, Janke AL, Collins DL, Baillet S. Brain templates and atlases. *NeuroImage*. 2012 Aug; 62(2):911–922.  
453 [doi:10.1016/j.neuroimage.2012.01.024](https://doi.org/10.1016/j.neuroimage.2012.01.024).
- 454 **Fischl B**, Sereno MI, Dale AM. Cortical surface-based analysis II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*. 1999; 9(2):195–207.  
455
- 456 **Fonov V**, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*. 2011 Jan; 54(1):313–327. [doi:10.1016/j.neuroimage.2010.07.033](https://doi.org/10.1016/j.neuroimage.2010.07.033).  
457
- 458 **Friston KJ**, Ashburner J, Kiebel SJ, Nichols TE, Penny WD. *Statistical parametric mapping : the analysis of functional brain images*. London: Academic Press; 2006.  
459
- 460 **Goerzen D**, Fowler C, Devenyi GA, Germann J, Madularu D, Chakravarty MM, Near J. An MRI-Derived Neuroanatomical Atlas of the Fischer 344 Rat Brain. *Sci Rep*. 2020 Apr; 10(1):6952. [doi:10.1038/s41598-020-63965-x](https://doi.org/10.1038/s41598-020-63965-x).  
461  
462
- 463 **Good CD**, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*. 2001; 14(1):21–36. [doi:10.1006/nimg.2001.0786](https://doi.org/10.1006/nimg.2001.0786).  
464
- 465 **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, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data*. 2016 Jun; 3:160044. [doi:10.1038/sdata.2016.44](https://doi.org/10.1038/sdata.2016.44).  
466  
467  
468
- 469 **Halchenko YO**, FSL user mailing list: Incorrect probabilities in Harvard-Oxford-sub Left hemisphere; 2013. [Online] Available: <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=FSL;2bb44bee.1301> (Accessed: 2021-07-05).  
470
- 471 **Halchenko YO**, Meyer K, Poldrack B, Solanky DS, Wagner AS, Gors J, MacFarlane D, Pustina D, Sochat V, Ghosh SS, Mönch C, Markiewicz CJ, Waite L, Shlyakhter I, Vega Adl, Hayashi S, Häusler CO, Poline JB, Kadelka T, Skytén K, et al. DataLad: distributed system for joint management of code, data, and their relationship. *Journal of Open Source Software*. 2021 Jul; 6(63):3262. [doi:10.21105/joss.03262](https://doi.org/10.21105/joss.03262).  
472  
473  
474
- 475 **Holmes CJ**, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR Images Using Registration for Signal Averaging. *J Comput Assist Tomogr*. 1998 Mar; 22(2):324–333. <https://insights.ovid.com/crossref?an=00004728-199803000-00032>.  
476  
477
- 478 **Jenkinson M**, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *NeuroImage*. 2012 Aug; 62(2):782–790.  
479 [doi:10.1016/j.neuroimage.2011.09.015](https://doi.org/10.1016/j.neuroimage.2011.09.015).
- 480 **Kjønigsen LJ**, Lillehaug S, Bjaalie JG, Witter MP, Leergaard TB. Waxholm Space atlas of the rat brain hippocampal region: Three-dimensional delineations based on magnetic resonance and diffusion tensor imaging. *NeuroImage*. 2015 Mar; 108:441–449. [doi:10.1016/j.neuroimage.2014.12.080](https://doi.org/10.1016/j.neuroimage.2014.12.080).  
481  
482
- 483 **Landman BA**, Ribbens A, Lucas B, Davatzikos C, Avants B, Ledig C, Ma D, Rueckert D, Vandermeulen D, Maes F. MICCAI 2012 Workshop on Multi-Atlas Labeling. CreateSpace Independent Publishing Platform; 2012.  
484
- 485 **Lüders E**, Steinmetz H, Jäncke L. Brain size and grey matter volume in the healthy human brain. *NeuroReport*. 2002 Dec; 13(17):2371–2374. [https://journals.lww.com/neuroreport/Abstract/2002/12030/Brain\\_size\\_and\\_grey\\_matter\\_volume\\_in\\_the\\_healthy.40.aspx](https://journals.lww.com/neuroreport/Abstract/2002/12030/Brain_size_and_grey_matter_volume_in_the_healthy.40.aspx).  
486  
487
- 488 **MacNicol E**, Ciric R, Kim E, Censo DD, Cash D, Poldrack R, Esteban O. Atlas-based brain extraction is robust across rat MRI studies. In: *IEEE 19th International Symposium on Biomedical Imaging (ISBI 2021)* Nice, France; 2021. p. (accepted). [doi:10.1109/ISBI48211.2021.9](https://doi.org/10.1109/ISBI48211.2021.9).  
489  
490
- 491 **Marcus DS**, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *J Cogn Neurosci*. 2007 Aug; 19(9):1498–1507. [doi:10.1162/jocn.2007.19.9.1498](https://doi.org/10.1162/jocn.2007.19.9.1498).  
492  
493
- 494 **Mars RB**, Passingham RE, Jbabdi S. Connectivity Fingerprints: From Areal Descriptions to Abstract Spaces. *Trends Cogn Sci*. 2018 Nov; 22(11):1026–1037. [doi:10.1016/j.tics.2018.08.009](https://doi.org/10.1016/j.tics.2018.08.009).  
495
- 496 **Martin RF**, Bowden DM. *Primate Brain Maps: Structure of the Macaque Brain*. Elsevier; 2000.
- 497 **Matsuzawa J**, Matsui M, Konishi T, Noguchi K, Gur RC, Bilker W, Miyawaki T. Age-related Volumetric Changes of Brain Gray and White Matter in Healthy Infants and Children. *Cereb Cortex*. 2001 Apr; 11(4):335–342.  
498 [doi:10.1093/cercor/11.4.335](https://doi.org/10.1093/cercor/11.4.335).  
499

- 500 **Mazziotta J**, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins  
501 L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons  
502 L, et al. A Four-Dimensional Probabilistic Atlas of the Human Brain. *J Am Med Inform Assoc.* 2001 Sep;  
503 8(5):401–430. doi:10.1136/jamia.2001.0080401.
- 504 **Mazziotta JC**, Toga AW, Evans A, Fox P, Lancaster J. A Probabilistic Atlas of the Human Brain: Theory and  
505 Rationale for Its Development: The International Consortium for Brain Mapping (ICBM). *NeuroImage.* 1995  
506 Jun; 2(2, Part A):89–101. doi:10.1006/nimg.1995.1012.
- 507 **MNI**, NeuroImaging & Surgical Technologies Lab (Montreal Neurological Institute) – MNI Atlases; n.d. [Online]  
508 Available: <http://nist.mni.mcgill.ca/category/atlas/> (Accessed: 2021-07-01).
- 509 **Nooner KB**, Colcombe S, Tobe R, Mennes M, Benedict M, Moreno A, Panek L, Brown S, Zavitz S, Li Q, Sikka S,  
510 Gutman D, Bangaru S, Schlachter RT, Kamiel S, Anwar A, Hinz C, Kaplan M, Rachlin A, Adelsberg S, et al. The  
511 NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. *Front Neurosci.*  
512 2012; 6. doi:10.3389/fnins.2012.00152.
- 513 **Osen KK**, Imad J, Wennberg AE, Papp EA, Leergaard TB. Waxholm Space atlas of the rat brain auditory sys-  
514 tem: Three-dimensional delineations based on structural and diffusion tensor magnetic resonance imaging.  
515 *NeuroImage.* 2019 Oct; 199:38–56. doi:10.1016/j.neuroimage.2019.05.016.
- 516 **Papp EA**, Leergaard TB, Calabrese E, Allan Johnson G, Bjaalie JG. Addendum to “Waxholm Space atlas of  
517 the Sprague Dawley rat brain” [*NeuroImage* 97 (2014) 374–386]. *NeuroImage.* 2015 Jan; 105:561–562.  
518 doi:10.1016/j.neuroimage.2014.10.017.
- 519 **Papp EA**, Leergaard TB, Calabrese E, Johnson GA, Bjaalie JG. Waxholm Space atlas of the Sprague Dawley rat  
520 brain. *NeuroImage.* 2014 Aug; 97:374–386. doi:10.1016/j.neuroimage.2014.04.001.
- 521 **Paxinos G**, Watson C. The rat brain in stereotaxic coordinates. Elsevier Academic Press; 1997.
- 522 **Pisner D**, Hammonds R. PyNets: A Reproducible Workflow for Structural and Functional Connectome Ensemble  
523 Learning. In: *Annual Meeting of the Organization for Human Brain Mapping*, vol. 26 Online Event; 2020. <https://github.com/dPys/PyNets/>.  
524
- 525 **Robinson EC**, Jbabdi S, Glasser MF, Andersson J, Burgess GC, Harms MP, Smith SM, Van Essen DC, Jenkinson  
526 M. MSM: A new flexible framework for Multimodal Surface Matching. *NeuroImage.* 2014; 100:414–426.  
527 doi:10.1016/j.neuroimage.2014.05.069.
- 528 **Rohlfing T**. Incorrect ICBM-DTI-81 atlas orientation and white matter labels. *Front Neurosci.* 2013; 7.  
529 doi:10.3389/fnins.2013.00004.
- 530 **Satterthwaite TD**, Connolly JJ, Ruparel K, Calkins ME, Jackson C, Elliott MA, Roalf DR, Hopson R, Prabhakaran  
531 K, Behr M, Qiu H, Mentch FD, Chiavacci R, Sleiman PMA, Gur RC, Hakonarson H, Gur RE. The Philadelphia  
532 Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain de-  
533 velopment in youth. *NeuroImage.* 2016 Jan; 124:1115–1119. doi:10.1016/j.neuroimage.2015.03.056.
- 534 **Sawiak SJ**, Wood NI, Williams GB, Morton AJ, Carpenter TA. SPMMouse: A new toolbox for SPM in the animal  
535 brain. In: *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 17 Hawaii, USA; 2009. p. 1086.
- 536 **Schaefer A**, Kong R, Gordon EM, Laumann TO, Zuo XN, Holmes AJ, Eickhoff SB, Yeo BTT. Local-Global Parcel-  
537 lation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex.* 2018 Sep;  
538 28(9):3095–3114. doi:10.1093/cercor/bhx179.
- 539 **Schurr PH**, Merrington WR. The Horsley–Clarke stereotaxic apparatus. *Br J Surg.* 1978; 65(1):33–36.  
540 doi:10.1002/bjs.1800650110.
- 541 **Shi F**, Yap PT, Wu G, Jia H, Gilmore JH, Lin W, Shen D. Infant Brain Atlases from Neonates to 1- and 2-Year-Olds.  
542 *PLoS One.* 2011 Apr; 6(4):1–11. doi:10.1371/journal.pone.0018746.
- 543 **Sowell ER**, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW. Mapping cortical change across  
544 the human life span. *Nat Neurosci.* 2003 Mar; 6(3):309–315. doi:10.1038/nn1008.
- 545 **Szulf KU**, Lerch JP, Nieman BJ, Bartelle BB, Friedel M, Suero-Abreu GA, Watson C, Joyner AL, Turnbull  
546 DH. 4D MEMRI atlas of neonatal FVB/N mouse brain development. *NeuroImage.* 2015 Sep; 118:49–62.  
547 doi:10.1016/j.neuroimage.2015.05.029.

- 548 **Talairach J**, Tournoux P. Co-planar stereotaxic atlas of the human brain. Stuttgart New York: Georg Thieme  
549 Verlag/Thieme Medical Publishers; 1988.
- 550 **Talairach J**, David M, Tournoux P, Corredor H, Kvasina J. Atlas d'anatomie stéréotaxique: repérage radiologique  
551 indirect des noyaux gris centraux des régions mésencéphalo-sous-optique et hypothalamique de l'homme.  
552 Masson; 1957.
- 553 **Tavor I**, Jones OP, Mars RB, Smith SM, Behrens TE, Jbabdi S. Task-free MRI predicts individual differences in  
554 brain activity during task performance. *Science*. 2016 Apr; 352(6282):216–220. doi:10.1126/science.aad8127.
- 555 **Thompson WH**, Fanton S, Open Source Software: NetPlotBrain; 2021. [Online] Available: [https://zenodo.org/  
556 record/4593837](https://zenodo.org/record/4593837) (Accessed: 2021-07-08).
- 557 **Tosun D**, Siddarth P, Levitt J, Caplan R. Cortical thickness and sulcal depth: insights on de-  
558 velopment and psychopathology in paediatric epilepsy. *BJPsych Open*. 2015 Oct; 1(2):129–135.  
559 doi:10.1192/bjpo.bp.115.001719.
- 560 **Van Essen DC**. Windows on the brain: the emerging role of atlases and databases in neuroscience. *Curr Opin*  
561 *Neurobiol*. 2002 Oct; 12(5):574–579. doi:10.1016/S0959-4388(02)00361-6.
- 562 **Van Essen DC**, Glasser MF, Dierker DL, Harwell J, Coalson T. Parcellations and Hemispheric Asymmetries of  
563 Human Cerebral Cortex Analyzed on Surface-Based Atlases. *Cerebral Cortex*. 2012 Oct; 22(10):2241–2262.  
564 doi:10.1093/cercor/bhr291.
- 565 **Von Economo CF**, Koskinas GN. Atlas of cytoarchitectonics of the adult human cerebral cortex, vol. 10. Karger  
566 Basel; 2008.
- 567 **Wilkinson MD**, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva San-  
568 tos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers  
569 R, Gonzalez-Beltran A, et al. The FAIR Guiding Principles for scientific data management and stewardship.  
570 *Sci Data*. 2016 Mar; 3(1):160018. doi:10.1038/sdata.2016.18.
- 571 **Yarkoni T**, Markiewicz C, de la Vega A, Gorgolewski K, Salo T, Halchenko Y, McNamara Q, DeStasio K, Po-  
572 line JB, Petrov D, Hayot-Sasson V, Nielson D, Carlin J, Kiar G, Whitaker K, DuPre E, Wagner A, Tirrell L,  
573 Jas M, Hanke M, et al. PyBIDS: Python tools for BIDS datasets. *J Open Source Softw*. 2019 Aug; 4:1294.  
574 doi:10.21105/joss.01294.
- 575 **Yeo BT**, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Poli-  
576 meni JR, Fischl B, Liu H, Buckner RL. The organization of the human cerebral cortex estimated by intrinsic  
577 functional connectivity. *J Neurophysiol*. 2011 Jun; 106(3):1125–1165. doi:10.1152/jn.00338.2011.
- 578 **Yoon U**, Fonov VS, Perusse D, Evans AC. The effect of template choice on morphometric analysis of pediatric  
579 brain data. *NeuroImage*. 2009 Apr; 45(3):769–777. doi:10.1016/j.neuroimage.2008.12.046.

## 580 **Online methods**

### 581 **Design and architecture**

582 *TemplateFlow* comprises four cardinal components: (i) a cloud-based archive, (ii) a Python client for  
583 programmatically querying the archive, (iii) automated systems for synchronizing and updating  
584 archive data, and (iv) inter-template registration workflows. Here, we discuss the details of each  
585 component's implementation in turn, as well as the manner in which they interact with one another  
586 to form a cohesive whole.

#### 587 *The TemplateFlow Archive.*

588 The archive itself comprises directories of template data in cloud storage. For redundancy, the  
589 data are stored on both Google Cloud using the Open Science Framework (OSF) and on Amazon's  
590 Simple Storage Service (S3). Prior to storage, all template data must be named and organized in  
591 directories conforming to a data structure adapted from the Brain Imaging Data Structure (BIDS)  
592 standard (*Gorgolewski et al., 2016*). The precise implementation of this data structure is a living  
593 document and is detailed on the *TemplateFlow* homepage (<http://www.templateflow.org>). We detail  
594 several critical features here.

595 The archive is organized hierarchically, and descriptive metadata follow a principle of inher-  
596 itance: any metadata that apply to a particular level of the archive also apply to all deeper lev-  
597 els. At the top level of the hierarchy are directories corresponding to each archived template. If  
598 applicable, within each template directory are directories corresponding to sub-cohort templates.  
599 Names of directories and resource files constitute a hierarchically ordered series of key-value pairs  
600 terminated by a suffix denoting the datatype. For instance, `tpl-MNIPediatricAsym_cohort-3_res-`  
601 `high_T1w.nii.gz` denotes a T1-weighted template image file for resolution "high" of cohort "3" in the  
602 "MNIPediatricAsym" template (where the definitions of each resolution and cohort are specified in  
603 the template metadata file, *TemplateFlow Archive*). The most common *TemplateFlow* datatypes are  
604 indexed in Suppl. Table S2; an exhaustive list is available in the most current version of the BIDS  
605 standard (<https://bids.neuroimaging.io/>).

606 Within each directory, template resources include image data, atlas and template metadata,  
607 transform files, licenses, and curation scripts. All image data are stored in gzipped NIFTI-1 format  
608 and are conformed to RAS+ orientation (i.e., left-to-right, posterior-to-anterior, inferior-to-superior,  
609 with the affine qform and sform matrices corresponding to a cardinal basis scaled to the resolution  
610 of the image). Template metadata are stored in a JavaScript Object Notation (JSON) file called `tem-`  
611 `plate_description.json`; an overview of metadata specifications is provided in Figure 3 of the main  
612 manuscript. In brief, template metadata files contain general template metadata (e.g., authors and  
613 curators, references), cohort-specific metadata (e.g., ages of subjects included in each cohort), and  
614 resolution-specific metadata (e.g., dimensions of images associated with each resolution). Atlas  
615 metadata are often stored in TSV format and specify the region name corresponding to each atlas  
616 label. Transform files are stored in HDF5 format and are generated as a diffeomorphic composition  
617 of ITK-formatted transforms mapping between each pair of templates.

618 The archive has a number of client-facing access points to facilitate browsing of resources. Key  
619 among these is the archive browser on the *TemplateFlow* homepage, which indexes all archived  
620 resources and provides a means for researchers to take inventory of possible templates to use for  
621 their study.

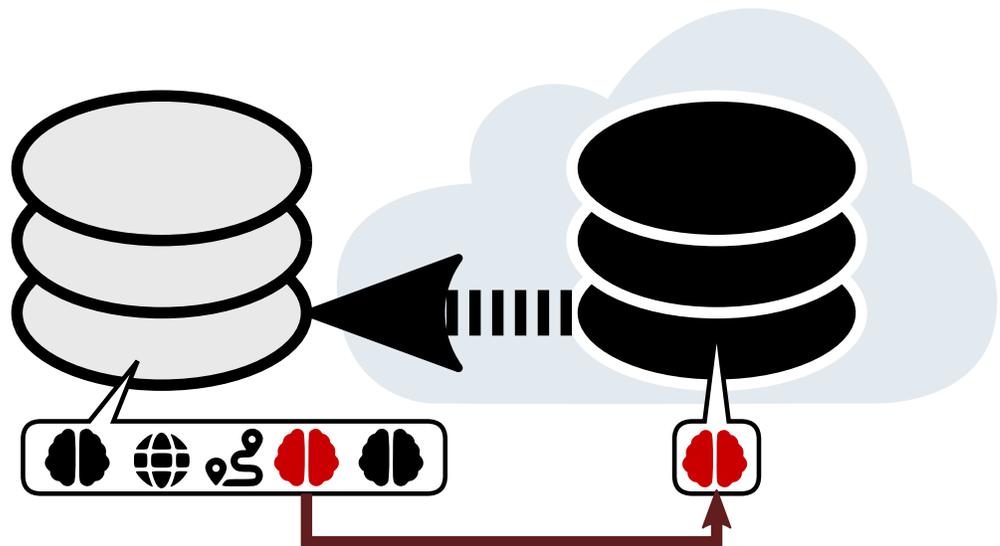
#### 622 *The Python client.*

623 *TemplateFlow* is distributed with a Python client that can submit queries to the archive and down-  
624 load any resources as they are requested by a user or program. Valid query options correspond  
625 approximately to BIDS key-value pairs and datatypes. A compendium of common query arguments  
626 is provided in Suppl. Table S2, and comprehensive documentation is available on the *TemplateFlow*  
627 homepage.

628 When a query is submitted to the *TemplateFlow* client, the client begins by identifying any files

**Figure 5. Example usage of the Python-based *TemplateFlow Client*.**

After importing the API, the user submits a query for the T1-weighted FSL version of the MNI template at 1 mm resolution. The client first filters through the archive, identifies any files that match the query, and finds their counterparts in cloud storage. It then downloads the requested files and returns their paths in the local *TemplateFlow* installation directory. Future queries for the same resource can be completed without any re-downloading.



```
>>> from templateflow import api as tflow
>>> tflow.get('MNI152NLin6Asym', desc=None, resolution=1,
...          suffix='T1w', extension='nii.gz')
PosixPath('/templateflow_home/tpl-MNI152NLin6Asym/tpl-
MNI152NLin6Asym_res-01_T1w.nii.gz')
```

629 in the archive that match the query. To do so, it uses *PyBIDS* (Yarkoni et al., 2019), which exploits  
630 the BIDS-like architecture of the *TemplateFlow Archive* to efficiently scan all directories and filter any  
631 matching files. Next, the client assesses whether queried files exist as data in local storage. When a  
632 user locally installs *TemplateFlow*, the local installation initially contains only lightweight pointers to  
633 files in OSF cloud storage. These pointers are implemented using *DataLad* (Halchenko et al., 2021),  
634 a data management tool that extends git and git-annex. *TemplateFlow* uses *DataLad* principally to  
635 synchronize datasets across machines and to perform version control by tracking updates made  
636 to a dataset.

637 If the queried files are not yet synchronized locally (i.e., they exist only as pointers to their  
638 counterparts in the cloud), the client instructs *DataLad* to retrieve them from cloud storage. In  
639 the event that *DataLad* fails or returns an error, the client falls back on redundancy in storage and  
640 downloads the file directly from Amazon's S3. When the client is next queried for the same file, it  
641 will detect that the file has already been cached in the local filesystem. The use of resource pointers  
642 with the client thus enables lazy loading of template resources. Finally, the client confirms that the  
643 file has been downloaded successfully. If the client detects a successful download, it returns the  
644 result of the query; in the event that it detects a synchronization failure, it displays a warning for  
645 each queried file that encountered a failure.

646 Continued functionality and operability of the client is ensured through an emphasis on maxi-  
647 mizing code coverage with unit tests. Updating the client requires successful completion of all unit  
648 tests, which are automatically executed by continuous integration (CI) and continuous delivery (CD)  
649 services connected to GitHub. CI and CD also keep the web-based archive browser up to date by  
650 automatically indexing data files.

651 Ancillary and managerial systems.

652 *TemplateFlow* includes a number of additional systems and programs that serve to automate stages  
653 of the archive update process, for instance addition of a new template or revision of current tem-  
654 plate resources. To facilitate the update and extension process, *TemplateFlow* uses GitHub actions

## Box 1. Integration of *TemplateFlow* in processing workflows

*TemplateFlow* maximizes the accessibility and reuse potential of templates and atlases. For example, let's reuse the base configuration file for FSL FEAT we proposed in our paper (*Esteban et al., 2019*). The design file `design.fsf` specifies a simple preprocessing workflow with FSL tools. The simplified code listing below shows that, just to make non-default templates available to FSL using the graphical user interface (GUI), at least four steps are necessary:

```
# 1. User determines two nondefault templates they want to spatially normalize into
# 2. User manually download templates, extract the required files from packages
$ curl -sSL <url> | tar zxv --no-same-owner -C /data/templates/
# 3. User opens FSL's GUI, edits the target template box content pointing to the appropriate files
# 4. User generates FSL configuration files to permit batch execution on the command line
# 5. For the default and the two nondefault templates, execute FSL's feat:
$ feat design_<template>.fsf
```

The outputs of each `feat design_<template>.fsf` call will follow the pre-specified patterns of FSL, with whatever customization the user has introduced into the design file. The user, therefore, must then adapt the downstream analysis tools to correctly interpret the derived dataset, in each standard space, or reformat the output dataset according to the expectations of the analysis tools.

The user is also responsible for all aspects of provenance tracking and adequately reporting them in their communications. Information such as version of the template (or download date), citations to relevant papers, and other metadata (e.g., RRDIDs) must be accounted for manually throughout the research process.

---

In contrast, tools using *TemplateFlow* dramatically simplify the whole process (note that `MNI152NLin2009cAsym` and `OASIS30Ants` are the two templates not found within the FSL distribution, and `MNI152NLin6Asym` denotes FSL's *MNI space* (i.e., the default FSL template):

```
$ fmriprep /data /derivatives participant --output-spaces MNI152NLin2009cAsym MNI152NLin6Asym OASIS30Ants
```

*fMRIPrep* generates the results with BIDS-Derivatives organization for the three templates. The tool also leverages *TemplateFlow* to generate a *boilerplate citation text* that includes the full names, versions and references to credit the template's authors for each of the templates involved.

*fMRIPrep* internally stages one spatial normalization workflow for each of the output spaces. Each of these normalization sub-workflows uses a simple line of Python code to retrieve the necessary resources from *TemplateFlow* using the *TemplateFlow Client* interface (Figure 5):

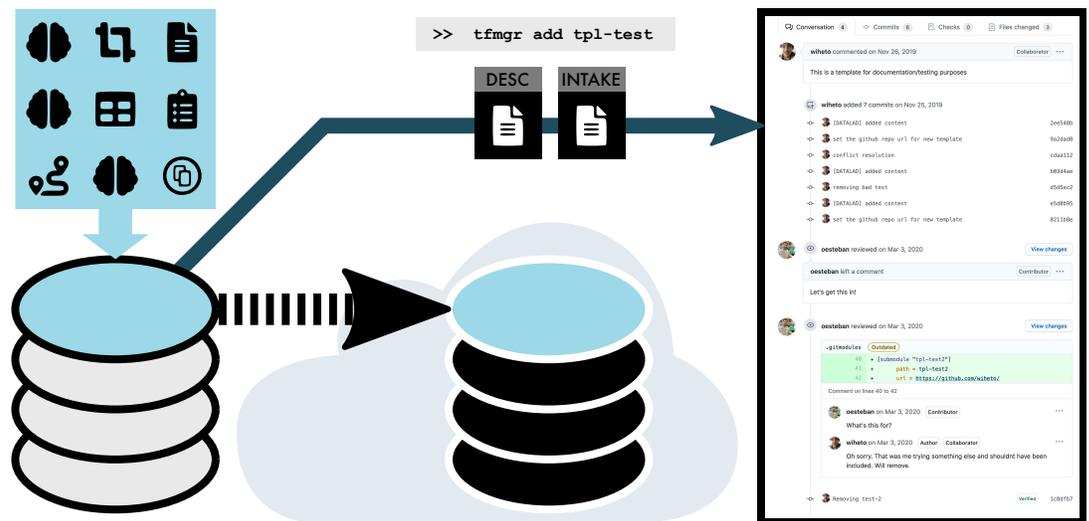
```
>>> from templateflow.api import get
>>> tpl_ref_file = get("MNI152NLin6Asym", desc=None, resolution=1, suffix="T1w", extension="nii.gz")
```

One detail overseen in the FSL example is that, for a robust spatial normalization process, a precise binary mask of the brain is generally used. While FSL would require the user to manually set this mask up in the GUI, in the case of *TemplateFlow*, it requires a second minimal call:

```
>>> msk_ref_file = get("MNI152NLin6Asym", desc="brain", resolution=1, suffix="mask", extension="nii.gz")
```

These examples are extreme simplifications of what a pipeline developer can automate and make more robust by integrating *TemplateFlow* in their workflows.

For further examples on how *TemplateFlow* can be leveraged, *PyNets* is a package for ensemble learning of functional and structural connectomes (*Pisner and Hammonds, 2020*), and *NetPlotBrain* for visualization (*Thompson and Fanton, 2021*).



**Figure 6. Contributing a new resource with the *TemplateFlow Manager*.** To contribute a new template to *TemplateFlow*, members of the community first organize template resources to conform to the BIDS-like *TemplateFlow* structure. Next, *tfmgr* (the *TemplateFlow Manager*, see Suppl. Table S3) synchronizes the resources to OSF cloud storage and opens a new pull request proposing the addition of the new template. A subsequent peer-review process ensures that all data are conformant with the *TemplateFlow* standard. Finally, *TemplateFlow* curators merge the pull request, thereby adding the template into the archive.

655 to automatically synchronize dataset information so that all references remain up-to-date with the  
 656 current dataset. These actions are triggered whenever a pull request to *TemplateFlow* is accepted.  
 657 For example, GitHub actions are used to update the browser of the *TemplateFlow Archive* so that it  
 658 displays all template resources as they are uploaded to the archive.

659 Whereas the *TemplateFlow Client* synchronizes data from cloud storage to the local filesystem,  
 660 a complementary *TemplateFlow Manager* handles the automated synchronization of data from the  
 661 local filesystem to cloud storage. The Python-based manager is also used for template intake, i.e.,  
 662 to propose the addition of new templates to the archive. To propose adding a new template, a user  
 663 first runs the *TemplateFlow Manager* using the `tfmgr add <template_id> --osf-project <project_id>`  
 664 command.

665 The manager begins by using the *TemplateFlow* client to query the archive and verify that the  
 666 proposed template does not already exist. After verifying that the proposed template is new, the  
 667 manager synchronizes all specified template resources to OSF cloud storage. It then creates a fork  
 668 of the `tpl-intake` branch of the *TemplateFlow* GitHub repository and generates an intake file in  
 669 Tom's Obvious Minimal Language (TOML) markup format; this intake file contains a reference to  
 670 the OSF project where the manager has stored template resources. The *TemplateFlow Manager*  
 671 commits the TOML intake file to the fork and pushes to the user's GitHub account. Finally, it re-  
 672 trieves template metadata from `template_description.json` and uses the metadata to compose  
 673 a pull request on the `tpl-intake` branch. This pull request provides a venue for discussion and  
 674 vetting of the proposed addition of a new template.

675 Inter-template registration workflow.

676 To enable the flow of knowledge across template spaces, *TemplateFlow* includes a workflow for com-  
 677 puting robust transformations between any pair of adult human template spaces. To compute a  
 678 transformation between two template spaces, the inter-template registration workflow makes use  
 679 of 10 of the high-quality T1-weighted adult human brain images used in the creation of the MNI  
 680 152 template portfolio. In the first step of the workflow, these 10 images are registered to both  
 681 template spaces using the symmetric normalization (SyN) algorithm (Avants et al., 2008). Next, a  
 682 10-channel registration is performed in ANTs using the SyN algorithm. Thus, the workflow com-

683 puts a single transformation that simultaneously optimizes the alignment between all 10 images  
684 in both coordinate spaces.

685 **MNI space text mining analysis.**

686 To investigate the use of the term “MNI” in the neuroimaging literature, we conducted an ex-  
687 ploratory text mining analysis. For this, we used the Elsevier API to download the entire corpus  
688 of two leading journals of neuroimaging methodology, *NeuroImage* and *NeuroImage: Clinical*. In  
689 this way, we retrieved a total 16,812 full-text articles that were subsequently segmented into lists  
690 of sentences. A scan of these sentences revealed 14,870 sentences across 6,048 articles that con-  
691 tained the word “MNI”. Sentences were cleaned (i.e., removing punctuation, single letters, accents,  
692 numbers) and tokenized into words, which were subsequently lemmatized (i.e., converted to base  
693 form) using the NLTK wordnet lemmatizer. From the lemmatized words, we filtered out stopwords  
694 (i.e., NLTK stopwords and a custom list) and included words with a frequency above 10 as part of  
695 our “dictionary”; this yielded a dictionary size of 2,324 words.

696 Next, we computed a sparse dictionary by article count matrix (i.e.,  $2,324 \times 6,048$ ), on which we  
697 performed topic modelling with latent Dirichlet allocation (LDA; *Blei et al., 2003*, implementation  
698 from scikit-learn with the learning decay hyperparameter set to 0.7). The number of topics ( $k=15$ )  
699 was selected by identifying the LDA model yielding the lowest perplexity (*Blei et al., 2003*) sweeping  
700 the interval [8-16] for the parameter. The 20 words from the dictionary that loaded the highest on  
701 the 15 topics were visualized using word clouds. Topics were sorted by descending dominance,  
702 and the dominance fraction (number of articles where the topic is the most loaded with respect  
703 to the total 6,048 documents) was included above the corresponding topic’s word cloud (Figure 4,  
704 and Suppl. Figure S1).

## Supp. Box S1. The FAIR Guiding Principles

(Reproduced from *Wilkinson et al., 2016*, Box 2)

### To be **Findable**:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

### To be **Accessible**:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
  - A1.1. the protocol is open, free, and universally implementable
  - A1.2. the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

### To be **Interoperable**:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

### To be **Reusable**:

- R1. (meta)data are richly described with a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards

## Supp. Box S2. Quick start with the *TemplateFlow Client API*

### Finding templates

At the time of writing there are 15 templates available within the resource, and all the unique identifiers can be accessed with the `templates()` method:

```
>>> from templateflow.api import templates
>>> f"{' '.join(templates())} ({len(templates())} templates)"
"""Fischer344, MNI152Lin, MNI152Nlin2009cAsym, MNI152Nlin2009cSym, MNI152Nlin6Asym, MNI152Nlin6Sym,
MNIInfant, MNIPediatricAsym, NKI, OASIS30ANTs, PNC, UNCInfant, WHS, fsLR, fsaverage (15 templates)"""
```

### Accessing Metadata

We can query metadata associated to individual data files (e.g., a volume or a surface) or general metadata of the template. For example, the `get_metadata(<template_id>)` returns the general metadata as a dictionary. Hence, consulting the full name corresponding to some template identifiers yields:

```
>>> from templateflow.api import get_metadata
>>> [(tpl, get_metadata(tpl)["Name"]) for tpl in templates()[:3]]
[("Fischer344", "MRI-Derived Neuroanatomical Atlas of the Fischer 344 Rat Brain"),
 ("MNI152Lin", "Linear ICBM Average Brain (ICBM152) Stereotaxic Registration Model"),
 ("MNI152Nlin2009cAsym", "ICBM 152 Nonlinear Asymmetrical template version 2009c")]
```

Similarly, we can check the license of a given template:

```
>>> get_metadata("UNCInfant")["License"]
"CC-BY"
```

Or the proper citations (please note that the output in this example has been manipulated for demonstration purposes):

```
>>> print(get_citations("UNCInfant", bibtex=True)[0])
@article{uncinfant1,
  doi = {10.1371/journal.pone.0018746},
  year = 2011,
  volume = {6},
  number = {4},
  pages = {e18746},
  author = {Feng Shi and Pew-Thian Yap and Guorong Wu and Hongjun Jia and John H. Gilmore and
    Weili Lin and Dinggang Shen},
  title = {Infant Brain Atlases from Neonates to 1- and 2-Year-Olds},
  journal = {{PLoS ONE} {ONE}}
}
```

```
>>> from templateflow.api import get
>>> print(get("MNI152Nlin6Asym", desc=None, resolution=1, suffix="T1w"))
Downloading https://templateflow.s3.amazonaws.com/tp1-MNI152Nlin6Asym/tp1-MNI152Nlin6Asym_res-01_T1w.nii.gz
100%|#####| 3.29M/3.29M [00:01<00:00, 2.11MB/s]
/home/oesteban/.cache/templateflow/tp1-MNI152Nlin6Asym/tp1-MNI152Nlin6Asym_res-01_T1w.nii.gz
```

**Supp. Table S1.** Common terms and definitions.

---

Atlas	A map between spatial coordinates and knowledge-based annotations, such as neuroanatomical structures or functional brain networks.
Template	A digital image that is representative of a population of interest (for instance an average of individual images from the population) and can be used as a spatial reference for that population.
Volumes and surfaces	Two common schemes for spatially organising brain data. Volumetric data are situated in the Euclidean topology of physical three-dimensional space, while surface data make use of the native topology of the cortex's sulcal curvature.
Stereotaxic coordinate system	A three-dimensional coordinate system that can be referenced to unambiguously localise a position in the brain, defined using an origin together with a basis set of axes.
Standard space	A coordinate system engendered by a template, which can be used as a reference when studying a population. A volumetric <i>template</i> engenders a <i>stereotaxic coordinate system</i> , which can be annotated in different ways to define brain <i>atlases</i> .
Transform	A diffeomorphic function that maps a set of spatial coordinates to their analogues or homologues in another coordinate space. For example, transforms can map from structures in the brains of individual subjects to analogous structures in a population-average template.

---

**Supp. Table S2.** *TemplateFlow* data entities.

Data entity	API query example	Description
Template	<code>"MNI152Lin"</code>	The template dataset to which an image or other data file belongs.
Resolution	<code>res=1</code>	The image resolution. Each resolution is assigned a key, which is defined in the <code>res</code> field of <code>template_description.json</code> .
Mask	<code>desc="brain", suffix="mask"</code>	Indicates that the image is a binary-valued annotation, where voxels labelled 1 are part of the mask.
Discrete segmentation	<code>desc="malf", suffix="dseg"</code>	Indicates that the image is an integer-valued annotation. Each segmentation image file ( <code>.nii.gz</code> format) is paired with a dictionary of segment names ( <code>.tsv</code> format).
Probabilistic segmentation	<code>label="CSF", suffix="probseg"</code>	Indicates that the image is a probabilistic annotation, wherein the value of each voxel indicates the probability of that voxel belonging to the specified label.
Atlas	<code>atlas="Schaefer", desc="7Network"</code>	The atlas to which a segmentation file belongs.
Transformation	<code>from="MNI152Lin", suffix="xfm"</code>	File containing a mapping between 2 stereotaxic coordinate spaces. The source space is defined in the <code>from</code> field, while the target space is defined in the <code>tpl</code> field.
Image modality	<code>suffix="T1w"</code>	For non-annotation brain images, the suffix indicates whether the image is T1-weighted ( <code>T1w</code> ), T2-weighted ( <code>T2w</code> ), proton density-weighted ( <code>PD</code> ), or T2*-weighted ( <code>T2star</code> ).
Template cohort	<code>cohort=1</code>	Subsample of a dataset used to generate an average template.

**Supp. Table S3.** Command-line interface for *TemplateFlow Manager*.

Argument	Environment variable	Specifications
<code>template_id</code>		Identifier of the template. This is the value of the <code>tpl</code> field in all file names.
<code>--osf-project</code>	<code>OSF_PROJECT</code>	The OSF project where the template data are to be stored. The project must be writable by the user account whose credentials are specified in the <code>--osf-user</code> and <code>--osf-password</code> arguments.
<code>--osf-user</code>	<code>OSF_USERNAME</code>	Account username or identifier for OSF cloud storage.
<code>--osf-password</code>	<code>OSF_PASSWORD</code>	Account password for OSF cloud storage.
<code>--osf-overwrite</code>		Flag that indicates that the OSF client should force the overwrite of any existing files in the OSF project that have names conflicting with those of new files.
<code>--gh-user</code>	<code>GITHUB_USER</code>	Account username for GitHub. The user account whose credentials are provided must have a fork of the TemplateFlow repo.
<code>--gh-password</code>	<code>GITHUB_PASSWORD</code>	Account password for GitHub.
<code>--path</code>	<code>OSF_PROJECT</code>	Path to a local directory where template resources are located. The path must either be a directory whose name is <code>tpl-&lt;template_id&gt;</code> or contain such a directory.
<code>--nprocs</code>		Maximum number of parallel processes to run when uploading to or fetching from OSF.



**Supp. Figure S1. Full report for the topic modeling exploration.** All 15 topics modeled are represented —sorted by dominance— in this figure, completing the partial results of Figure 4.