



# Place-Making Narrative Data: Management Issues in the Context of Open Science and Data Curation in France

**WILLIAM KELLEHER**

**HILLARY BAYS**

*\*Author affiliations can be found in the back matter of this article*

RESEARCH PAPER

]u[ubiquity press

## ABSTRACT

The Places dataset is still in-process and being added to on a regular basis. It is comprised of complete data management documentation and ethics documentation in English and French, concept notes in English and French, and participant data files that include .csv metadata sheets, .wav audio recordings of interviews, .jpeg photographs of the place of the interview and open ELAN .eaf transcriptions. The participant interview data relies on a narrative inquiry methodology and is collected in one-on-one interviews with researchers on the theme of anglophone expatriation and adaptation to French socio-physical space. These interviews are informal and interactive, aiming for a sharing of experience, since the researchers are, themselves, anglophone expatriates. The aim of this accompanying article is to provide information on the dataset through a concomitant reflection on the research context in France and the affordances of the increasingly powerful TGIR Huma-Num digital humanities infrastructure, as well as the tensions that are prevalent in the Franco-European space between the demands of Open Science and the restrictions of the European General Data Protection Regulations. These affordances and this tension have framed the process that has resulted in the Places data architecture.

CORRESPONDING AUTHOR:

**William Kelleher**

LIDILE, Université Rennes 2,  
Rennes, Bretagne, France

[william.kelleher@univ-rennes2.fr](mailto:william.kelleher@univ-rennes2.fr)

KEYWORDS:

open data; data management plan (DMP); dataset; TGIR Huma-Num; ELAN; transcription; narrative; expatriation

TO CITE THIS ARTICLE:

Kelleher, W., & Bays, H. (2022). Place-Making Narrative Data: Management Issues in the Context of Open Science and Data Curation in France. *Journal of Open Humanities Data*, 8: 12, pp. 1–13. DOI: <https://doi.org/10.5334/johd.75>

## (1) INTRODUCTION

Places is an interview-based narrative research project lead by Hillary Bays and William Kelleher from the LISAA research unit at Gustave Eiffel and the LIDILE research unit at Rennes 2, respectively.

Places data is opened on the Nakala data repository that is overseen by the Digital Humanities Very Large Research Infrastructure (Sciences Humaines Numériques Très Grande Infrastructure de Recherche – TGIR Huma-Num) (CNRS, 2022). The project and data collection are ongoing. The Nakala data set includes full data management documentation, full ethics documentation in English and French, concept notes in English and French and participant data files that include .csv metadata sheets, .wav audio recordings of interviews, .jpeg photographs of the place of the interview and open ELAN (MPI, 2021) .eaf transcriptions.

The Places project seeks to better understand processes of place-making, primarily in response to the dramatic shifts in socialisation, language, and cultural adequation undergone by those in situations of expatriation in France. These shifts suppose, additionally, a reflection on events such as the COVID pandemic, new configurations of work, and the merging of the offline and the online. Places participants are anglophone expatriates in France. Expatriates, generally, are critically aware of their surrounds because of their need to adapt to new circumstances. Places is compiling an open dataset of narratives. In these narratives, material urban fabric becomes the interface for biographic work on displacement – whether linguistic, personal, or professional.

Whereas the social context for Places is an interrogation of public space, the wider research context is the gathering force of the move towards Open Science, and the adoption of that move in France, and Franco-European space. Open science is a dynamic new movement among French universities, and Places is very much the result of the evolving resources, directives and support available. This paper will present some key elements of the French framework for Open Science and discuss the Places data and how it fits into this new and exciting research landscape.

## (2) OPEN SCIENCE IN FRANCE

### (2.1) INSTITUTIONAL SUPPORT

Open Science in France is prompted, amongst other factors, by a perceived need both to publish more in English-speaking journals and to change the direction of research (from top-down to bottom-up) This change would make research accessible to associations and NGO's, which, in France, are known as the 'third' sector ('tiers' secteur). Open Science initiatives adhere to FAIR guidelines (Wilkinson et al, 2016) in the sense of open research protocols and data storage, but also in terms of accessible and interoperable source codes and machine-to-machine communication. Overarching guidance is provided by the Open Science Committee (Comité pour la science ouverte – CSO<sup>1</sup>) of the Ministry for Higher Education, Research and Innovation (Ministre de l'Enseignement Supérieur, de la Recherche et de l'Innovation – MESRI<sup>2</sup>) which responds to European Commission recommendation 2018/790 of 25 April 2018 (CSO, 2018). The second national implementation plan was recently published (MESRI, 2021).

The Open Science Committee and the MESRI dialogue with two principal centres. The Centre for Direct Scientific Communication (Centre pour la Communication Scientifique Directe – CCSD<sup>3</sup>) is responsible for the HAL<sup>4</sup> platforms which curate open research outputs. The Institute for Scientific and Technical Information (Institut de l'information scientifique et technique – Inist<sup>5</sup>) governs two bodies promoting best practice. Research Data: Digital Learning (Données de la Recherche: Apprentissage numérique – DoRANum<sup>6</sup>) and Optimising Sharing and Interoperability

---

1 <https://www.ouvrirlascience.fr/accueil/> (last accessed: 16 March 2022).

2 <https://www.enseignementsup-recherche.gouv.fr/fr> (last accessed: 16 March 2022).

3 <https://www.ccsd.cnrs.fr/> (last accessed: 16 March 2022).

4 <https://hal.archives-ouvertes.fr/> (last accessed: 16 March 2022).

5 <https://www.inist.fr/> (last accessed: 16 March 2022).

6 <https://doranum.fr/> (last accessed: 16 March 2022).

of Research Data (Optimiser le Partage et l'Interopérabilité des Données de la Recherche – OPIDoR<sup>7</sup>) that applies the Science Europe<sup>8</sup> guidelines and templates.

A very central Open Science initiative is the Digital Humanities Very Large Research Infrastructure (Sciences Humaines Numériques Très Grande Infrastructure de Recherche – TGIR Huma-Num). As its name would indicate, it aims to be a complete research infrastructure. One can connect with a specific HumanID<sup>9</sup> or through a series of professional and academic logons (eduGAIN,<sup>10</sup> ORCID,<sup>11</sup> HAL, LinkedIn,<sup>12</sup> Twitter<sup>13</sup> and Google<sup>14</sup>). The TGIR Huma-Num landing page opens out onto a series of platforms. Places makes use of Nakala and Sharedocs. Nakala is a dataset managing and hosting repository that is specifically built for scientific data. It offers assignation of a Digital Object Identifier (DOI<sup>15</sup>), hierarchisation (using 'collections'), description, metadata and a selection of Creative Commons sharing licences.

TGIR Huma-Num is a central partner in a series of research initiatives and systems, all of which promote Open Science. The Common Language Resources and Technology Infrastructure (CLARIN<sup>16</sup>) is one example that is very important. CLARIN is a European label that hosts and curates larger corpuses generated by research consortiums. TGIR Huma-Num also supports smaller data-sharing platforms. Ortolang<sup>17</sup> and COCOON<sup>18</sup> are examples of the latter. Ortolang is a corpus repository for linguistics that hosts phonetic annotations, multimedia files, and multimodal or text-based datasets. TGIR Huma-Num promotes Open Science in other ways, for instance through the free availability of a host of textometric and other analytic tools, like TXM.<sup>19</sup>

French universities are very active partners in open research that is conceived of as pluri-site and localisable. Open Science charters (see Université Rennes 2 a, 2021) or initiatives such as Keys for Open Science (Clés pour la Science Ouverte – SOcle<sup>20</sup>) (Université Rennes 2 b, 2022), for example, serve as references for training and information. Infrastructures such as TGIR Huma-Num, similarly have contacts that can be reached directly by telephone or email, thanks to the Human Sciences Foundation (Fondation Maison des Sciences de l'Homme – MSH<sup>21</sup>).

The Nakala repository that has been adopted as the Places data hosting service is a very good example of Open Science in France. It is dialogable with other TGIR Huma-Num applications and is a fun and accessible platform, as attested to by its furry-eared logo given in **Extract 1**:



**Extract 1** The TGIR Huma-Num Nakala data repository logo.

7 <https://opidor.fr/> (last accessed: 16 March 2022).

8 <https://www.scienceeurope.org/about-us/> (last accessed: 16 March 2022).

9 <https://humanid.huma-num.fr/> (last accessed: 16 March 2022).

10 <https://discovery.renater.fr/edugain/> (last accessed: 16 March 2022).

11 <https://orcid.org/> (last accessed: 16 March 2022).

12 <https://fr.linkedin.com> (last accessed: 16 March 2022).

13 <https://www.twitter.com> (last accessed: 16 March 2022).

14 <https://www.google.com> (last accessed: 16 March 2022).

15 <https://www.doi.org/> (last accessed: 16 March 2022).

16 <https://www.clarin.eu/> (last accessed: 16 March 2022).

17 <https://www.ortolang.fr/> (last accessed: 16 March 2022).

18 <https://cocoon.huma-num.fr/exist/crdo/> (last accessed: 16 March 2022).

19 <https://txm.gitpages.huma-num.fr/textometrie/> (last accessed: 16 March 2022).

20 One will note, in passing, that French acronyms often have an inbuilt play on words. 'Socle', for instance, translates as hub, or pedestal.

21 <https://www.fmsch.fr/en> (last accessed: 16 March 2022).

In many ways, Nakala is similar to other open repositories such as figshare.<sup>22</sup> Its main difference is that it is provided with more purview by TGIR Huma-Num. Nakala is not specifically orientated towards linguistics, unlike Ortolang. It hosts data from anthropology and history, etc. and promotes transversality in research. Although this is perhaps a point that has already implicitly been made, Nakala is, refreshingly in this age of GAFA, a non-for-profit, publicly-funded service.

## **(2.2) OPEN SCIENCE AND CURATORIAL ISSUES IN THE FRANCO-EUROPEAN SPACE**

Open Science initiatives like TGIR Huma-Num, whose objective can be defined as “transparent and accessible knowledge that is shared and developed through collaborative networks” (Vicente-Saez & Martinez-Fuentes, 2018: 434), enter into tension with the demands of access control and confidentiality set out, in the Franco-European space, by the European General Data Protection Regulation – GDPR 2016/679 that was adopted in France as Law 0141 of 21 June 2018 (JORF, 2018).

The 2018 European General Data Protection Regulation (Le règlement général sur la protection des données – RGPD) severely restricts what types of data can be collected and how they can be stored. The RGPD forms part of the National Commission for Digital Rights and Freedoms (Commission Nationale de l’Informatique et des Libertés – CNIL<sup>23</sup>). It is much more disciplinary than the Helsinki declaration (World Medical Association, 2013). It requires a data protection officer for each research institution and an acknowledgement of the bases of the research. One of three possible bases may be chosen: i) public interest, ii) legitimate interest, or iii) legal obligation.

Project organisation is affected by the RGPD, since a deadline must be set for destruction of those documents (such as consent forms, or primary documentation) that identify or implicate participants and both sensitive and confidential information must be removed:

- Sensitive data includes racial or ethnic references, political opinions, religious and philosophical convictions, trade union affiliation, genetic and biometric data, health data, sexual data, penal record, identification document data and numbers.
- Confidential data includes names and addresses of persons directly related to the participant, proper names such as residential street addresses, telephone numbers, work names and addresses and specific identifying times and events.

Application of the RGPD has led to the establishment of university ethics committees that are, in some ways, a new part of the research environment in France. Even in 2008, the French academic community was still discussing whether ethics committees were even a positive addition (see Vassy & Keller, 2008). Prior to ethics committees, universities had primarily engaged in considerations of copyright and image reproduction. The RGPD is, therefore, also part of a wide-ranging change in the European research landscape, reflected in documentation (see Liège Université, 2022) that is advancing a participant-centred research approach.

Perhaps there are differing points of departure for Open Science and for the RGPD. Whereas the RGPD emphasises protection, Open Science emphasises divulgation. The Nakala repository is a good example here as well. To paraphrase its documentation (TGIR Huma-Num, 2022), there are three basic settings for uploaded data: i) uploaded, unpublished, and private, ii) published and public, iii) published and embargoed. Options i) and iii) are not open, option ii) is completely open. It is therefore hard to find a middle term that is suitable for qualitative data. Often, only institutional repositories offer made-to-measure safeguards that can include traceability or agree and close licences, such as those applied to the Talking Fish corpus (Goodall, Frawley & Sewell, 2011).

It is in the Data Management Plans (DMP’s) for projects, such as Places, that these tensions between data openness and data protection play out. A DMP must negotiate the twin imperatives of curatorial fireguards (see also Bouchet-Moneret, 2021; Stérin, 2018) and the growing demand for open data. In the following sections, we will look more closely at the Places project, its theoretical framework, and methods before considering curatorial decisions.

---

<sup>22</sup> <https://figshare.com/> (last accessed: 16 March 2022).

<sup>23</sup> <https://www.cnil.fr/> (last accessed: 16 March 2022).

### (3) THE PLACES RESEARCH PROJECT

#### (3.1) THEORETICAL FRAMEWORK

The theoretical framework of Places is narrative inquiry, as this term is understood by Bamberg (Bamberg & Demuth, 2016), and by the small stories movement (Bamberg & Georgakopoulou, 2008; De Fina & Georgakopoulou, 2008; Georgakopoulou, 2007; Patron, 2020). As participants engage in storytelling, they accomplish specific identity work. Places focuses on this work as a means of understanding both linguistic behaviour and orientation to broader societal discourses and processes. Narratives are both a means of portraying our conception of our world to others and a means of making that world comprehensible to ourselves. The employment of time (series of events) and space (the arena in which those events occur) results in discursive units (small stories) that are transferrable and that may be borrowed, lent, expanded upon, or alluded to.

Narrative inquiry encourages the creation of projects that seek to unravel narrative space and to mobilise narrative analysis in order to better understand lived processes of change and adaptation. A very good example of such processes is expatriation, when a participant leaves what is familiar, familial, and close in order to move to a foreign country in search of, or to take up, work. Small story discourse units are part of material space since they assist us in our 'making' of place and locality (Appadurai, 1996). This is especially pertinent in these COVID-informed times that simultaneously combine proxemic isolation, hyper connectedness, and fragility (see Fine & Abramson, 2020).

#### (3.2) PARTICIPANTS AND DATA

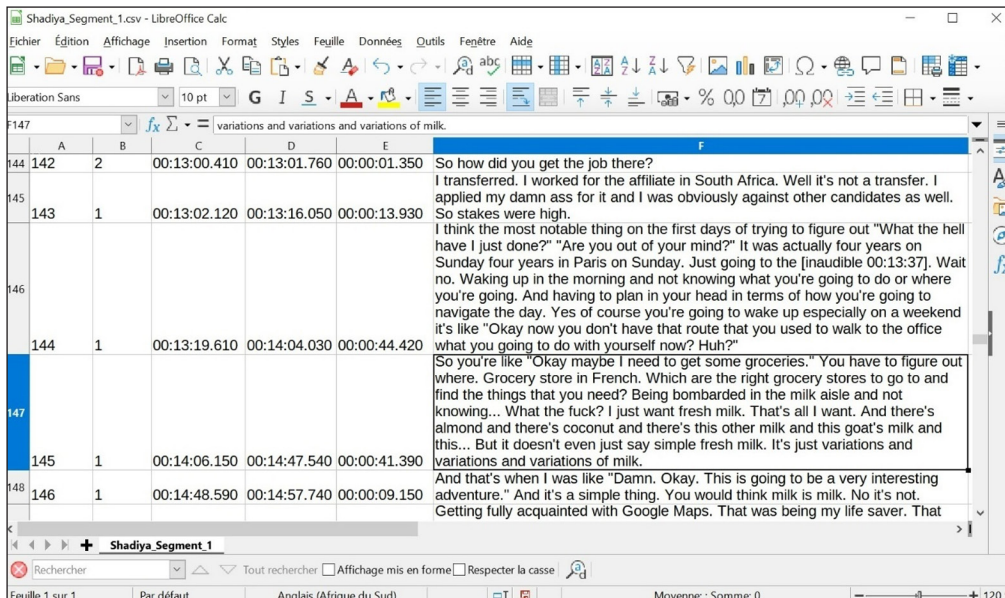
Places is a project in process which aims to assemble 40 participants, or approximately 20hrs of fully transcribed talk. Places recruits English-language participants, or participants from predominantly English-speaking countries, who have been expatriated to France. Expatriation is taken simply to mean residing in a foreign country for reasons that involve work, as opposed to humanitarian or migratory reasons.<sup>24</sup> Expatriates maintain links to their home country and often return frequently. Participants are recruited using the 'snowball' method that involves contacts through associations or through networks of acquaintance. After making contact with the researchers, participants are invited to an interview. It is an informal interview format that is envisaged as a sharing between researcher and participant, in an (auto-) biographical fashion, since both lead researchers are also expatriates. A guide to possible questions and themes raised is provided in the DMP. Interviews take place in spaces frequented, and chosen, by the participants.

The researcher and participant meet there. The researcher takes photographs of the site (with a high-definition Canon EOS200D for example) and sets up the recording device (a tripod-mounted Zoom H2n). The advantage of this kind of recording device is that it provides a 2- or 4-way surround that incorporates the noises of the site. It is a portable device so that, should the participant decide to conduct the researcher around the site, it is possible to continue recording. Outside interviewing in a place that is of affective significance to a participant stimulates a strong degree of narrativization, of the emplotting, in narratives, of that place. As Georgakopoulou puts it, "I view every space as place: as an experienced, lived, and practised social arena by social actors." (Georgakopoulou, 2015, p.65). Outside interviewing also mitigates against the 'Sunday' conditions that are often applicable to interview formats (see Bamberg, 2016, p. 17). In other words, outside interviewing confronts storytelling with the places and the material conditions to which it refers. A personal relationship to lived space is exemplified in *Extract 2* which is from an interview with a participant whose pseudonym is Shadiya.

Shadiya evokes how even the simple, day-to-day gestures that she had known (in South Africa) become difficult and strange. Although the type of shop, a supermarket, is the same as what she is used to, its layout and the products it sells in France are unknown.

---

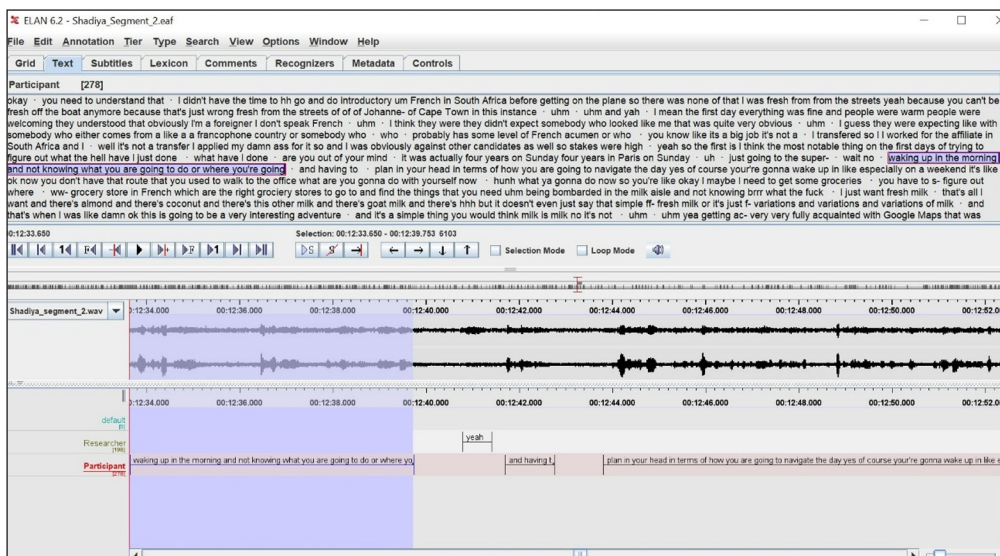
<sup>24</sup> The interview with the participant Declan includes a very interesting reflection on expatriation at the end.



**Extract 2** Participant Shadiya talks about grocery shopping in France.

### (3.3) METHODS FOR DATA ANALYSIS

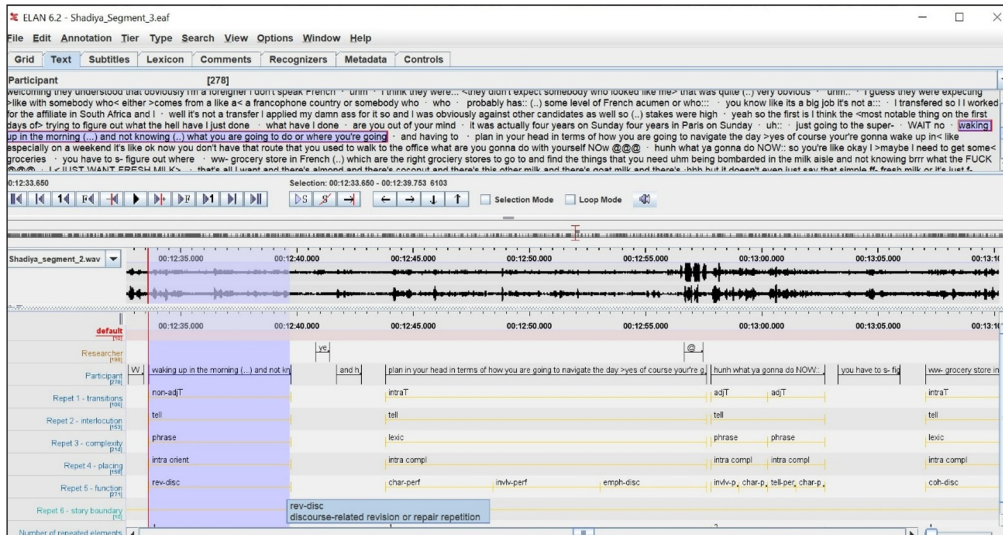
The .csv format in **Extract 2** transcribes the interview following editing for sensitive and confidential information. This first verbatim transcription is then corrected and segmented by the researchers, working in collaboration on Sharedocs, in order to produce an ELAN .eaf transcription (given in **Extract 3**) that is timed down to intonation groups and that includes markers of backchannelling etc. It is this second transcription that is verified by the researchers which will be uploaded to Nakala.



**Extract 3** The same segment as in Extract 2 that is now marked against the .wav waveform and separated by intonation groups.

One can note that this version of the transcription has much shorter annotation blocks and indicates backchannelling etc. It is also a much closer transcription in that hesitations, false starts, and repairs are also transcribed. It is a transcription with which data analysis can begin. ELAN offers several advantages. It marks annotations directly onto a .wav audio file and does so with a discrimination of 1000<sup>th</sup> of a second. It offers a series of tiers for different participants and for different kinds of annotation. A research project can create a customised ELAN template that uses controlled vocabularies to standardise annotations between different researchers.

A controlled vocabulary is a series of terms of interest to a particular type of analysis that can be pre-defined for a specific tier. When a segment of waveform is selected for annotation, the annotator chooses among the pre-defined terms This leads to a greater regularity in annotation. As a result, annotations can be compared across different dataset items. An example of tiers and controlled vocabularies is given in **Extract 4**.



**Extract 4** The same segment as in Extract 2 that has now been coded for instances of repetition, using controlled vocabularies, and tiers.

**Extract 4** is an example of a first research orientation of Places that investigates repetition. The different tiers concern: i) the dynamics of repetition between interlocutors and between ‘turns’ or intonation groups, ii) the complexity of the repetition, iii) the placing of the repetition within story units or conversational projects, and iv) the discourse-related or performance-related function of the repetition (a distinction taken from Auer, 1984). ELAN easily exports .xml or .csv formats that allow transcriptions to be easily embedded in other applications, and, thanks to the accuracy of the annotations on the waveform, other kinds of analysis, like phonemics using tools such as Praat (Boersma & Weenink, 2022), can be applied.

## (4) PLACES AS AN EXAMPLE OF OPEN SCIENCE IN FRANCE

Places, in its conception, development and, now, data collection and compilation, is very much interlinked with Open Science initiatives in France. Its Data Management Plan was developed with the aid of the OPIDoR<sup>25</sup> platform. The narratives it collects will be hosted on the Nakala repository. Its ethics clearance process has proceeded through one of the recently established committees, and its data architecture generally has benefitted from the advice and support of personnel involved either in research units, Open Science platforms, TGIR Huma-Num or the Human Sciences Foundation. A discussion of its data management, ethics, and reuse is a good way of exemplifying Open Science as it has been adopted in France. All documentation referred to below has been made available in the Places dataset (Kelleher & Bays, 2022) hosted on Nakala.

### (4.1) DATA MANAGEMENT

Since more than one researcher is involved in Places, its Data Management Plan has several key aims: i) to stabilise the interview process, ii) to standardise the project storage architecture (file extensions etc), iii) to increase the reliability and comparability of ELAN transcriptions uploaded to Nakala (their annotation), and iv) to ensure the regularity and findability of the dataset itself.

As concerns the first point, i) the interview process is supported by comprehensive project information sheets, a guide to possible questions and themes. Participants are made aware of the aims and rationale of the project through concept notes, the DMP, and through this paper and the Places dataset itself. One of the aims of pursuing an open and shared architecture is to offer future participants the possibility to be able to listen to and read other testimonies and biographies gathered in the context of the project, thereby building a cumulative archive. A data management checklist ensures that the same steps are followed for each participant: a) first contact, b) choice of interview site, c) recording and photographing of the interview, d) data storage on the researcher PC, e) sharing and inter-researcher editing on Sharedocs, and f) uploading on Nakala.

<sup>25</sup> <https://opidor.fr/> (last accessed: 16 March 2022).

The second point, ii) raises questions of project storage architecture. Storage architecture concerns the hierarchisation of files and formats, their naming, and the handling of access, rights, and procedures for revising and updating. Places drew up its data plan in line with DoRANum best practice (DoRANum, 2022) and has made available file naming conventions. The architecture aims to ensure that Places data is archived and treated in a comparable way across researchers. It is also a means of ensuring ethical compliance since it avoids stray files through numerical progression in pre-edited, edited, shared, and hosted file versions.

Point three, iii) linguistic and critical discursive analysis are normed activities, falling under ISO 12620:2019 that sets standards for analytic and pragmatic terms and categories. This allows ELAN controlled vocabularies to be marked to standards, should one so wish, and for metadata to be interoperable across ELAN and Nakala. ELAN, in a similar fashion, allows a great deal of annotator and participant information to be allocated to each tier. Since Places is generating a considerable amount of transcription work, the decision was made to use a third-party service for each initial transcription. HappyScribe<sup>26</sup> was involved in the project and, after consultation with their engineer, now offers an ELAN-importable .csv format for their transcriptions. Before sending to HappyScribe, as noted, each researcher edits the waveform for sensitive and confidential information (see discussion above). Use of a transcription service allows an impartial and quality-controlled base from which to start more detailed transcription (see section 3.4 above).

Point four, iv) the Nakala repository uses, typically, the Dublin Core Metadata Initiative<sup>27</sup> for collections and datasets. For each participant data file, a .csv metadata sheet compiled during the interview includes the non-invasive information that a participant is comfortable with sharing and that will assist in the discoverability and, therefore, reuse of the data. One of the reasons for the choice of Nakala is that it partakes in a national drive<sup>28</sup> towards perennialisation of data.

#### **(4.2) ETHICAL CONSIDERATIONS**

Places data is pseudonymised rather than anonymised. This is to say that whilst the participant names are changed, and there is no confidential or sensitive identifying information retained in the Nakala dataset. Interviews are concerned with a participant's personal appreciations and life experiences. This kind of data remains much more identifiable than data collected in the context of large institutional research where one participant's response is equivalent to those responses given by other members of the cohort. This made the finding of a middle ground, between completely open and completely locked data (see discussion in section 2.2 above) so much more important. This is to say that Places data should be open and available, but it should not be something that can be downloaded irresponsibly or automatically by data trawlers.

The DMP that accompanies this article is the fifth iteration of the document. This shows how long the consultative process has been to achieve a satisfactory solution to this ethical conundrum. A first, tentative, solution was to upload the dataset and complete its metadata so that it would be findable, but to embargo the participant data subject to written request made to the lead researchers. This proved to be unsustainable for several reasons: it required too much of institutional data officers, it was not consultable by the participants, and, to a certain extent, did little more than place the problem at one remove, since any data administration raised the same questions of allowable and disallowable access. The second solution, presented in this article, has been to upload all project data and documentation in corresponding files that are password zipped using 7-Zip.

This solution seems to be satisfactory in light of the following: i) it allows Places data to be entirely accessible to participants and researchers, ii) it allows Nakala to serve its primary objective as an open, consultable, project archive where one can receive updates on data itself and publications related to the data, iii) actual use of the data requires the downloading of the full participant data file with its associated metadata, and iv) since the password is

---

26 <https://www.happyscribe.com/> (last accessed: 16 March 2022).

27 <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/> (last accessed: 16 March 2022).

28 See <https://www.archivistes.org/> (last accessed: 16 March 2022).



given in the ReadMe which also lists full project information, the contacts of the researchers, the institutional data officers, and the applicable Creative Commons licence, access to the password implies acceptance of the ReadMe contents, v) finally, 7-Zip technology provides a level of protection against trawlers.

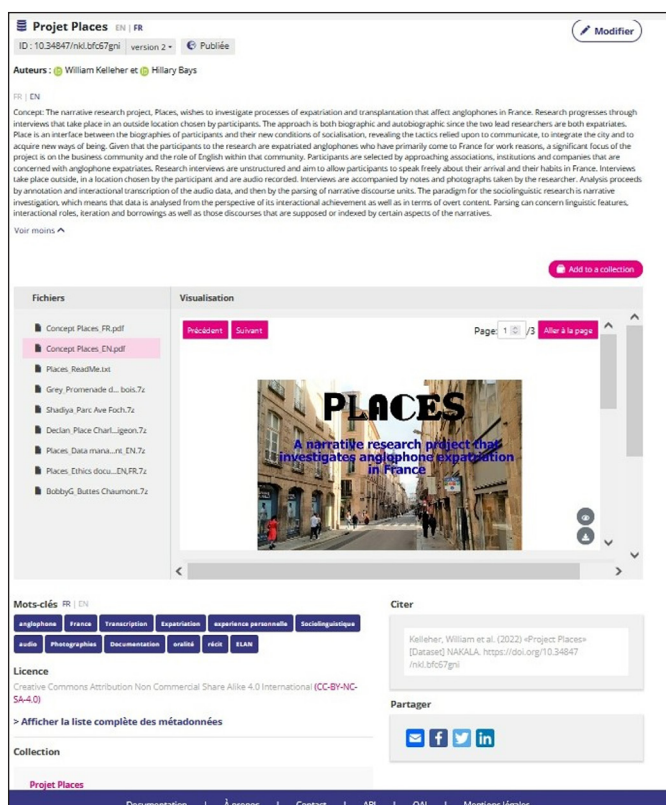
Generally, in its ethical response, Places has implemented the so-called 'data and paper' rule, where data is stored on disk whilst identifying information is signed and completed on paper. Paper can be stored in a designated, locked, space, and much more straightforwardly destroyed at the end of the data collection phase of the project. In particular with regards to security questions of hacking or surveillance capitalism, Places is doing the right thing by separating the paper trace and electronic data storage. It is a response to the new impetuses of open data and ethics movements in France.

### (4.3) APPLICATIONS AND RE-USE

Re-use of the Places dataset (Kelleher & Bays, 2022) is two-fold. On the one hand, there is a proposed re-use of the project documentation and data architecture. On the other, there is the data itself and its re-use for both linguistic and socio-historic purposes.

Concerning re-use of the project documentation and data architecture, several elements can be isolated: i) the use of the repository, metadata, and storage architecture, ii) data management plans and documentation, and, finally, iii) ethics documentation. Re-use of these elements would allow a humanities project to more easily transition to a research population in, or related to, francophone countries, since they offer bilingual French/English translations. They are also illustrative of an European research environment generally.

The solution that has been found for storage of different participant data elements in zipped files, associated with DCMI metadata and stored under a collection, has been achieved in consultation with different actors associated with open data and with the TGIR Huma-Num infrastructure. It is a solution that can be recycled in different projects. It allows sizeable data stockage, the association of individual .csv metadata files with each participant, and the assemblage of different kinds of data: audio, photographs, and transcription files. It is adapted to the Nakala platform that, hopefully, will gain in popularity and use. Individually password protecting the 7-Zip files and providing the password in the ReadMe contiguous with project information and Creative Commons share-alike licence protects the data from casual access and from data trawlers. A screen capture of the Nakala repository for Places, and the contents of a participant data file are given in *Extracts 5* and *6*.



Extract 5 A screen capture of the Nakala data presentation.

Nom	Taille	Compres...	Modifié le	Attributs	CRC	Chiffrer	Méthode	Bloc	Dossiers	Fichiers
BobbyG_Photos	6 199 384	0	2022-03-...	D	80239CB3	-			0	1
BobbyG_Buttes Chaumont.eaf	228 956	5 340 496	2022-03-...	A	3DD05168	+	LZMA2:2...	0		
BobbyG_Buttes Chaumont.wav	260 114 ...	151 074 ...	2022-03-...	A	8D25E628	+	Delta:4 L...	1		
BobbyG_Metadata sheet.csv	1 060		2022-03-...	A	880E68C7	+	LZMA2:2...	0		

**Extract 6** The contents of a 7-Zip participant data file.

The complete data management plans and associated documentation, such as file extension guidelines, project progress checklist, and metadata template represent best practice in the Franco-European research context. They have been included so that any project that involves European participants can have an example of this documentation. The same is true of the ethics documentation, its formulation, and specific references to Data Protection Officer and access conditions by participants and other researchers. The examples of information notices, consent forms, registry information, and interview guides are also intended as examples of best practice and have been provided in duplicate with a version in English and a version in French. This is a template that can, again, be used by another project.

One of the additional uses of a Nakala-hosted dataset concerns the participants themselves. Even if they do not consult their data, they are aware that they are contributing to an archive of contemporary speech and language behaviour. This further provides an element of reflexivity to Places. As researchers work on the data, they are conscious of the fact that it will be made public and invite comment and criticism. Open dataset hosting is part of the action-research paradigm, as evoked by the University of Liege ethics committee (2022). In this respect, one use of the Nakala platform is to serve in communication between the project leaders and the participants. Nakala hosts the data, but it can also indicate links to all those publications and conferences that use the data by including a DOI on the data page. This means that participants can track and be informed of what's happening to their data. This is, perhaps, one of the nicest ancillary benefits of the repository. Once the data are published, a link can be shared with those involved in the project.

The Places dataset is designed to promote continuing investigation. The .eaf transcription files are fully modulable, which means that a person accessing the data can reannotate on ELAN, export to other textometric applications such as *Atlas.ti* (Scientific Software Development GmbH, 2022), AntConc (Anthony, 2022) or TGIR Huma-Num's TXM<sup>29</sup>, dialogue with other platforms and reintegrate data into new corpuses. These corpuses can be place-based, since, in each case, one of the metadata entries is the site of the interview. One very significant goal of Places is, in this light, to contribute to the valuing of spoken narratives and to better understand how public space is accessed in our present conjuncture. The tier-based functioning of ELAN means that the dataset can be studied for patterning across entries, since a tier can focus on a specific aspect of analysis (prosody, morpho-syntax, semantics, paralinguistics, etc), be extracted, and then compared to a similar tier in a different entry.

Re-use of participant data can therefore concern studies of place (useful for ethnographic analysis), studies of interaction, and the features and functions of linguistic realisation. The participants are also members of a very specific, and quite rare, cohort. They are English speakers in a situation of enforced and prolonged language contact. This corpus, then, can also be used in studies of language variation. BobbyG's data is a good example of trilingual contact, as an example.

<sup>29</sup> <https://txm.gitpages.huma-num.fr/textometrie/> (last accessed: 16 March 2022).

## (5) CONCLUSIONS

The above sections have given the research methodology for Places and the choices informing the collection and hosting of its dataset on the Nakala repository. This has raised questions of Open Science in France and the tools available to accompany Data Management Plans and for facilitating research.

This short paper aims to contribute to Open Science. It has given links to all the major institutions and frameworks in France and provided a brief introduction to the research context. It has provided links to the Places dataset (Kelleher & Bays, 2022) on Nakala, its full documentation, and DMP. It has discussed legal and ethical frameworks for research in France and in Europe. Finally, it has explained the researchers' approach to data analysis. The ELAN .eaf files are hosted on Nakala (see Kelleher & Bays, 2022). They can be consulted, re-annotated, and re-shared in accordance with Creative Commons 4.0 International (CC-BY-NC-SA-4.0).

## DISCLAIMER

This article does not represent the views of either the researchers' host universities, research units, or the research project participants.

## DATA ACCESSIBILITY STATEMENT

The data to which this article refers is hosted on the Nakala repository with DOI: <https://doi.org/10.34847/nkl.bfc67gni>.

## ACKNOWLEDGEMENTS

The researchers would like to thank the participants to the project and the support received for Places from their respective research units, LIDILE and LISAA.

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR CONTRIBUTIONS

**William Kelleher:** conceptualisation, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing – original draft.

**Hillary Bays:** conceptualisation, funding acquisition, investigation, methodology, resources, validation, visualization, writing – review & editing.

## AUTHOR AFFILIATIONS

**William Kelleher**  [orcid.org/0000-0001-8368-6404](https://orcid.org/0000-0001-8368-6404)  
LIDILE, Université Rennes 2, Rennes, Bretagne, France

**Hillary Bays**  [orcid.org/0000-0001-5984-7431](https://orcid.org/0000-0001-5984-7431)  
LISAA, Université Gustave Eiffel, Paris, France

## REFERENCES

- Anthony, L.** (2022). *AntConc: a freeware corpus analysis toolkit for concordancing and text analysis*. Retrieved from <https://www.laurenceanthony.net/software/antconcl/> (last accessed: 16 March 2022).
- Appadurai, A.** (1996). *Modernity at large: cultural dimensions of globalisation*. University of Minnesota Press.
- Auer, J. C. P.** (1984). *Bilingual conversation*. John Benjamins. DOI: <https://doi.org/10.1075/pb.v.8>
- Bamberg, M., & Demuth, C.** (2016). Narrative inquiry: an interview with Michael Bamberg. *EJOP – Europe's Journal of Psychology*, 12(1), 14–28. DOI: <https://doi.org/10.5964/ejop.v12i1.1128>
- Bamberg, M., & Georgakopoulou, A.** (2008). Small stories as a new perspective in narrative and identity analysis. *Text & Talk*, 28(3), 377–396. DOI: <https://doi.org/10.1515/TEXT.2008.018>

- Boersma, P., & Weenink, D.** (2022). *Praat: doing phonetics by computer*. University of Amsterdam. Retrieved from <https://www.fon.hum.uva.nl/praat/> (last accessed: 16 March 2022).
- Bouchet-Moneret, F.** (2021). Les données personnelles de recherche et le RGPD. *Maison des Sciences de l'Homme de Lorraine*. Retrieved from <https://msh-lorraine.fr/wp-content/uploads/2022/01/BIS-Guide-Donnees-personnelles-et-RGPD.docx> (last accessed: 16 March 2022).
- Centre national de la recherche scientifique – CNRS.** (2022). *Huma-Num Très Grande Infrastructure de Recherche*. Retrieved from <https://www.huma-num.fr/> (last accessed: 16 March 2022).
- Comité pour la science ouverte.** (2018). Recommandation (UE) 2018/790 de la Commission du 25 avril 2018. Retrieved from <https://www.ouvrirlascience.fr/recommandation-ue-2018-790-de-la-commission-du-25-avril-2018-relative-a-laces-aux-informations-scientifiques-et-a-leur-conservation/> (last accessed: 16 March 2022).
- De Fina, A., & Georgakopoulou, A.** (2008). Analysing narratives as practices. *Qualitative Research*, 8(3), 379–387. DOI: <https://doi.org/10.1177/1468794106093634>
- Données de la Recherche: Apprentissage Numérique – DoRAnum.** (2022). *Comment bien nommer ses fichiers* (How to best name one's files). Retrieved from: <https://doranum.fr/stockage-archivage/comment-nommer-fichiers/> (last accessed: 16 March 2022).
- Fine, G. A., & Abramson, C. M.** (2020). Ethnography in the time of COVID-19: vectors and the vulnerable. *Etnografia e ricerca qualitativa*. July(2), 165–174. DOI: <https://doi.org/10.3240/97802>
- Georgakopoulou, A.** (2007). *Small stories, interaction and identities*. John Benjamins Publishing. DOI: <https://doi.org/10.1075/sin.8>
- Georgakopoulou, A.** (2015). Sharing as rescripting: place manipulations on YouTube between narrative and social media affordances. *Discourse, Context & Media*, 9, 64–72. DOI: <https://doi.org/10.1016/j.dcm.2015.07.002>
- Goodall, H., Frawley, J., & Sewell, H.** (2011). *Talking Fish*. dharmae: University of Technology Sydney. <https://dharmae.research.uts.edu.au/collections/show/24>
- Journal de la République Française – JORF.** (2018). LOI n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles. Retrieved from <https://www.legifrance.gouv.fr/jorf/jo/2018/06/21/0141> (last accessed: 16 March 2022).
- Kelleher, W., & Bays, H.** (2022). *Projet Places* [Dataset]. Nakala. DOI: <https://doi.org/10.34847/nkl.bfc67gni>
- Liège Université.** (2022). *Comité d'éthique en sciences humaines et sociales*. Retrieved from [https://www.recherche.uliege.be/cms/c\\_11292422/fr/comite-d-ethique-en-sciences-humaines-et-sociales](https://www.recherche.uliege.be/cms/c_11292422/fr/comite-d-ethique-en-sciences-humaines-et-sociales) (last accessed: 16 March 2022).
- Max Planck Institute for Psycholinguistics – MPI.** (2021). ELAN (version 6.2). *The Language Archive*. Retrieved from <https://archive.mpi.nl/tla/elan> (last accessed: 16 March 2022).
- Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation – MESRI.** (2021). *Deuxième plan national pour la science ouverte: généraliser la science ouverte en France 2021–2024*. Retrieved from [https://www.ouvrirlascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte\\_2021-2024.pdf](https://www.ouvrirlascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte_2021-2024.pdf) (last accessed: 16 March 2022).
- Patron, S.** (Ed.). (2020). *Small stories: un nouveau paradigme pour les recherches sur le récit*. Hermann.
- Scientific Software Development GmbH.** (2022). *Atlas.ti*. Retrieved from <https://atlasti.com/> (last accessed: 16 March 2022).
- Stérim, A.-L.** (2018). *Le point sur les données à caractère personnel*. Retrieved from <https://ethiquedroit.hypotheses.org/1717> (last accessed: 16 March 2022).
- TGIR Huma-Num.** (2022). *Documentation Nakala*. Retrieved from <https://documentation.huma-num.fr/nakala/> (last accessed: 16 March 2022).
- Université Rennes 2 a.** (2021). *Proposition pour une charte d'engagement en faveur de la science ouverte*. Retrieved from [https://socle.univ-rennes2.fr/system/files/UHB/Charte%20SO%20UR2\\_2021.pdf](https://socle.univ-rennes2.fr/system/files/UHB/Charte%20SO%20UR2_2021.pdf) (last accessed: 16 March 2022).
- Université Rennes 2 b.** (2022). *Clés pour la Science Ouverte – SOcle (Keys for Open Science)*. Retrieved from <https://socle.univ-rennes2.fr/> (last accessed: 16 March 2022).
- Vassy, C., & Keller, R.** (2008). Faut-il contrôler les aspects éthiques de la recherche en sciences sociales, et comment? *La découverte 'mouvements'*, 3(55–56), 128–141. DOI: <https://doi.org/10.3917/mouv.055.0128>
- Vicente-Saez, R., & Martinez-Fuentes, C.** (2018). Open Science now: a systematic literature review for an integrated definition. *Journal of Business Research*, 88(2018), 428–436. DOI: <https://doi.org/10.1016/j.jbusres.2017.12.043>
- Wilkinson, M. D., et al.** (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(160018), 1–9. DOI: <https://doi.org/10.1038/sdata.2016.18>
- World Medical Association.** (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA – Journal of the American Medical Association*, 310(20), 2191–2194. DOI: <https://doi.org/10.1001/jama.2013.281053>

**TO CITE THIS ARTICLE:**

Kelleher, W., & Bays, H. (2022).  
Place-Making Narrative Data:  
Management Issues in the  
Context of Open Science  
and Data Curation in France.  
*Journal of Open Humanities  
Data*, 8: 12, pp. 1–13. DOI:  
<https://doi.org/10.5334/johd.75>

Published: 06 May 2022

**COPYRIGHT:**

© 2022 The Author(s). This is an  
open-access article distributed  
under the terms of the Creative  
Commons Attribution 4.0  
International License (CC-BY  
4.0), which permits unrestricted  
use, distribution, and  
reproduction in any medium,  
provided the original author  
and source are credited. See  
[http://creativecommons.org/  
licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

*Journal of Open Humanities  
Data* is a peer-reviewed open  
access journal published by  
Ubiquity Press.