

DUCCIO PICCARDI, SILVIA CALAMAI

Fear of FAIR? Towards a new Italian incentive to oral data curation¹

This paper deals with the climate of distrust permeating contemporary society and how it may affect the researchers' adherence to initiatives dealing with the promotion of data FAIRness. A case in point is the perplexing state of a recent Italian action, a new census of Tuscan oral archives (*Gra.fo Reloaded* project) in which the respondents were asked to disclose several details about the oral documents they gathered. As a potential alternative strategy to get the ball of data curation rolling, this paper explores the current utilitarian mindset adopted by major international linguistics communities and details its application to the conception of a novel Italian academic outlet: *Oral Archives Journal* (OAr), a journal on oral archive conservation, description, and use, as well as on speech transcription, technologies, legal issues, and research ethics.

Keywords: Oral archives, data journal, trust research, data FAIRness, Tuscan census.

1. “*To be fair, I don’t trust them*”. *Research data in a cynical age*

As the basis of cooperative endeavors, mutual trust is a pivotal component of any kind of society. Alarmingly, several reports underline a global trust shortage, painting the picture of a contemporary distrusting world (Rotenberg, 2018). Even though the multifaceted nature of trust and its several measurement strategies (Bauer, Freitag, 2018) might cast a shadow of doubt on the generalizability of this crisis scenario (Brehm, Savel, 2019), its persistence in discourse indexes a great deal of concern and a constant fostering of ideas regarding institutional corruption and untrustworthiness (Rotenberg, 2018). According to Hardin (2006), trust has gradually gained relevance as a consequence of major societal trends, such as the growing prevalence of loose-knit networks and political hurdles in efficiently dealing with economic and security crises. Technology does its part, as the chaotic nature of information on the web favors a post-truth milieu in which institutional and expert claims are often met with skepticism (e.g., Harsin, 2018). Be that as it may, a recent contribution by Citrin and Stoker (2018) dubbed our times “a cynical age”, in which the (political) actions of others are assumed to always have self-serving motives. Similar phenomena are described in other domains: for example, Lambin (2008) defined consumer “fear of marketing” as an insidious mistrust stemming

¹ This contribution is the result of the joint work of the two authors. Duccio Piccardi is responsible for writing the text. Silvia Calamai is responsible for supervising the text and the project *Gra.fo Reloaded*.

from the assumption of constantly being the target of manipulative strategies, whose inescapable threads are woven by firm experts.

In this context, the concept of openness emerged in twenty-first century science as an attempt to make research practices transparent and accessible, ultimately revamping trust (Stracke, 2020). Research data management is a core concern of open science. Most notably, the principles enucleated in Wilkinson et al. (2016) gained traction through their widespread adoption by public initiatives: among those, the European Research Infrastructure Consortium CLARIN defines itself as an “*avant la lettre*” promoter of keeping data Findable, Accessible, Interoperable and Reusable (FAIR; De Jong, Maegaard, Fišer, Van Uytvanck & Witt, 2020: 3406). Of course, requiring drastic changes in researchers’ mindsets, open science is not something that can be achieved instantly. Several factors explain noticeable asymmetries in the endorsement of its tenets. Individual disciplinary communities of practice faced the issue of openness in different times and ways: linguistics, for example, is considered a “late-comer to the scientific frontier of replicable research” (Janda, 2021: 447); thus, a recent survey underlined a very slow uptake of open science practices and an overall unsatisfactory current situation (Bochynska, Keeble, Halfacre, Casillas, Champagne, Chen, Röthlisberger, Buchanan & Roettger, 2023).

More intriguingly, research has shown that the propensity to adhere to specific scientific habits, such as data sharing, is related to contextual and individual factors (research group norms, researcher attitudes, perceived barriers, sex, age, academic rank, and geographic location; Bamberger, Reeves, 2021). Gomes, Pottier, Crystal-Ornelas, Hudgins, Foroughirad, Sánchez-Reyes, Turba, Martinez, Moreau, Bertram, Smout & Gaynor (2022) conducted a comprehensive survey of the reported reasons for open science non-compliance, which can be roughly divided in three types (rearranging the original taxonomy):

- a. Knowledge barriers with respect to data archiving, rights, and sensitiveness (these can be lowered through the activities of local research communities, as it will be shown below, § 3.1).
- b. Lack of incentives, since data archiving takes time and effort, but has no obvious reward (this aspect is being dealt with through recent strategies of academic acknowledgement, as it will be shown below, § 3.2.1).
- c. Trust issues. Sharing data makes researchers feel exposed to comments about their potential errors or flaws, which, in turn, may reduce others’ trust in them. Moreover, a common concern pertains to original data misinterpretation and inappropriate use. Lastly, researchers tend to avoid the risks of scooping, i.e., being anticipated by others in conducting planned analyses, thus losing publication opportunities. This habit is grounded on false assumptions: in reality, researchers tend to prioritize original data collection over data re-use. For this reason, Dijkers (2019: 175) was led to the blunt statement that “the fear of “research parasites” likely is pathologic”.

All in all, open science seems to be an enterprise aiming to turn a vicious circle of distrust in a virtuous circle of trust, but making the first move could be not as straightforward as it may appear.

On a positive note, investigating researchers' behavior as data donors (i.e., survey participants) Fichtner, Horstmeier, Brühmann, Watter, Binder & Knaus (2023) found that the explicit mention of data sharing does not scare them away from completing the requested task. In the remainder of this paper (§ 2), a different picture will be drawn, as a recent CLARIN-IT-powered initiative, i.e., a new Tuscan oral archive census (one of the activities planned in the *Gra.fo Reloaded* project) struggled to take off due to gargantuan nonresponse rates. Of course, context is the key to interpreting these discrepancies. Belliard, Mainieri, Plomp, Ramos Padilla, Sun & Jeddi (2023) recently synthesized ten procedural points to develop FAIR discussions in specific environments: noticeably, among those, the authors include identifying the features of the target community and establishing adequate rewards for the adoption of the principles. Here we argue that in Italy, where utilitarianism is a central component of the attitudes expressed towards European institutions (Conti, Memoli, 2015), we could try to overcome trust barriers in open science by developing appealing incentives. For this reason, we will try to advance a local instantiation of the international discussion on making data curation a research product per se through dedicated publication outlets, not without mentioning what the Italian community has already accomplished to lower the knowledge barriers surrounding these issues in its country (§ 3).

2. *The perplexing state of affairs of a new oral archive census*

Gra.fo Reloaded was a project on oral archive preservation and re-use run by the Italian Regione Toscana between June 2022 and the end of November 2023; it was coordinated by Siena University, with the participation of Fondazione Sistema Toscana, Soprintendenza Archivistica e Bibliografica della Toscana, Ecomuseo della Montagna Pistoiese, Istituto di Linguistica Computazionale "Antonio Zampolli" (Pisa National Research Council) and CLARIN-IT. As a first step towards devising effective preservation strategies, the project launched a new census of the oral archives gathered in Tuscany, aiming to both update previous seminal works (e.g., Andreini, Clemente, 2007) and create a research community revolving around novel research tools and initiatives, such as *Archivio Vi.Vo.*, a platform under construction for oral archive restoration, description, and access (Calamai, Piccardi, Pretto, Candeo, Stamuli & Monachini, 2022), and the *Vademecum* described in § 3.1, respectively.

After a pre-test phase (Piccardi, Calamai, 2023), the census was published on Google Forms on December 10th, 2022. At the beginning of the census, two layers of text tried to consolidate the participant trust in the initiative, underlining its community-oriented motives (Benson-Greenwald, Trujillo, White & Dickman, 2023), its compliance with CLARIN-IT, and the adherence to the FAIR principles

of the promoted tools. As the census was purely preliminary to any kind of intervention, by no means were participants asked to agree to rights transfer or oral data deposit or disclosure.

In a first phase, through a circular cover letter asking for individual completion and dissemination of the initiative, the census form was sent to 369 email addresses, including individual researchers (selected through the critical analysis of their publication history), Italian scientific associations (linguistics, oral/public history, anthropology, sociology, musicology), Ph.D. coordinators of Tuscan universities, the offices of all Tuscan municipalities, and a small selection of cultural foundations. Moreover, all the involved partners advertised the census through their platforms, and a related article was published in the local spin-off of the daily National newspaper *La Repubblica* (2023, February 11)². Lastly, the census was promoted during related in-person events.

In a second phase, which began towards the end of March 2023 and was formally halted after the end of the project, the census was divulged to other target individuals through personalized emails, phone calls, or social media accounts. In addition to the aforementioned categories, we also reached for representatives of U.S. Italian cultural foundations and ex-M.o.A./Ph.D. students who appeared (in institutional thesis repositories) to have gathered oral documents in Tuscany, for a total of more than fifty additional contacts.

After four months and shortly after the beginning of the second phase (mid-April 2023), we gathered 14 responses; after almost a year, at the end of the project, this number increased to a total of 23. As a rough approximation, leaving aside opaque point-to-mass diffusion mechanisms and potential technical issues in individual email sending or receiving, the whole census publication period led us to a whopping 95 % nonresponse rate. In oral archive census endeavors, difficulties in response collection are far from being unheard of. In the introductory remarks of the first Italian census of the public oral archives, Mulè (1993: 35-36) noted that only half of the previously pinpointed institutes replied to the questionnaire. A few years later, in a preliminary report on the state of oral archives in Ireland, Ferriter (1998: 91) observed that “many did not reply”. Nonetheless, our quantities might signal, so to say, an eloquent silence which should be adequately interpreted (Ephratt, 2008). Thus, we looked at both the objective (i.e., who has actually responded) and subjective (i.e., the perplexities and doubts expressed by cautious respondents through personal communications) elements at our disposal, ending up with a picture resembling the above-discussed three barriers against open science (cf. Gomes et al., 2022):

² <https://www.dfclam.unisi.it/sites/st08/files/allegatiparagrafo/13-02-2023/2023021153753080-1.pdf>. Accessed 08 October 2024.

- a. Silence as “I don’t know”. Respondents expressed sincere unawareness on crucial circumstantial information on the data they possess, such as ownership, copyright, and privacy issues, hindering their willingness to disclose details about their materials³.
- b. Silence as “it’s not (academically) worth it”. Admittedly, because of its extensiveness, completing our census did require a certain level of dedication and time. Of our 23 responses, only five were sent by full-time active academic personnel, the others being written by students, temporary and independent researchers, retired individuals. This disproportion highlights the unattractive nature of data curation for those individuals who are already busy with many other institutional tasks and responsibilities, calling for effective incentive development actions.
- c. Silence as “I don’t trust”. More than half of the respondents did have previous contacts and/or in-person encounters with one of the two authors of this paper. Perceived social closeness is a pivotal component of cooperativeness and trust (e.g., Brudner, Karousatos, Fareri & Delgado, 2022). However, in order to avoid circularity, an initiative striving for community building should not majorly rely on pre-existing relationships. Moreover, even though, leaving aside an informal requested declaration of interest for the *Archivio Vi.Vo.* archival platform, the census did not imply in any way the drafting of a data deposit agreement, we noticed that our form was often erroneously interpreted as such, triggering defensive behaviors (see above, § 1).

Overall, this perplexing state of affairs points to the necessity of urgently furthering the general awareness on data curation: in order to do so, researchers should be provided with the right knowledge toolkits and incentives.

3. “To be FAIR, I might study and publish with them”. A *Vademecum* and a journal

In the previous section, we highlighted the putative role played by three widely discussed barriers slowing down the adoption of an open-science mindset in the abysmal response rate of an Italian Regional census of oral archives. In this section, we will briefly address the crux of the matter: how can researchers bootstrap their knowledge about oral data curation? Why should they spend their time and efforts to organize their data? What are their actual benefits in joining the trusting circle of data FAIRness?

3.1 Knowledge barrier: The *Vademecum per il trattamento delle fonti orali*

The *Vademecum per il trattamento delle fonti orali* (Tavolo permanente per le fonti orali, 2023) is, at present, a five-year collaborative enterprise, aiming at

³ Indeed, according to the responses given to specific sections of our census, only one participant had some form of legal training before gathering her oral documents. Similarly, only two had some prior archival knowledge, while the majority of the respondents (13) were well aware of the technical aspects of recording (tools, audio formats, etc.). It seems evident that technical training alone cannot make up for the lack of expertise in other pertinent domains.

offering both the Italian scientific communities and non-academic stakeholders a bootstrap guide on the conservation, the description, the use, and the re-use of oral archives (Calamai, Casellato & Stamuli, 2022). Very different subjects from academia and from the main agencies of the Ministry of Cultural Heritage worked together with private and public institutions and CLARIN-IT in order to solve long-standing structural problems related to oral sources. After a collaborative process of co-creation and public revision, a set of good practices were defined in a cross-disciplinary approach with respect to production, management, access, and preservation of audio materials produced not only within research projects but also public history initiatives and citizen science data collection. A first digital version was finalized in 2021 and made available through the websites of all the involved institutions and the official pages of the work group representing the *Vademecum*⁴. In October, 2023, a printed version was released with a substantial new section on speech transcription.

As it was originally encouraged through the round table held at the XV AISV annual conference (Piccardi, Ardolino & Calamai, eds., 2019), the *Vademecum* undertaking has always been deeply connected with open discussion. At the present time, a yearly meeting is organized in order to showcase its latest developments and related projects; moreover, the *Vademecum* has been presented to Italian University students and the general public through dedicated lessons and other types of events. However, having a grasp of its actual reach is quite problematic, since works referencing its use are just beginning to be published (e.g., Manali, 2024; Clemente, 2023 for an example of Master of Arts thesis). Here we argue that probing the public awareness of the *Vademecum* existence through multiple methodologies (e.g., periodic literature reviews, items inserted in pertinent questionnaires, etc.) will be a necessary task in order to trigger word of mouth in specific communities and organize focused dissemination events. The *Gra.fo Reloaded* census did contain a section about the *Vademecum*. Crucially, the vast majority of the respondents (18) had never heard of it before filling in the census form, but was eager to know more through the subscription to the *Vademecum* newsletter (16).

3.2 Incentive barrier: the *Oral Archives Journal* (OAr)

3.2.1 Historical background

Being a freely available enchiridion to oral data curation, the *Vademecum* could be able to cover for the knowledge needs of the Italian community. Consequently, we started to envision a way to give more value to the efforts of local researchers to engage in this kind of activities by looking at relevant discussions emerging in the linguistics milieu. At the beginning of the 2010s, both Australian and American linguistic societies published resolutions (see Thieberger, 2012 for a recollection of these documents) to ensure the recognition of language documentation archives as scholarly contributions. A few years later, Haspelmath, Michaelis

⁴ <https://sites.google.com/view/tavolopermanenteperlefontioral/chi-siamo>. Accessed 08 October 2024.

(2014) envisioned a “corpus journal” hosting peer-reviewed annotated corpora in order to incentivize (quality) data curation through the standard mechanisms of academic career-building. Of course, at the time, the idea of similar outlets was already piloted in other research fields (see Thieberger, Margetts, Morey & Musgrave, 2016). In particular, Lawrence, Jones, Matthews, Pepler & Callaghan (2011) listed a taxonomy of data publication models including overlay data journals, i.e., academic outlets publishing data description documents through a peer-review process encompassing both the document and the actual data curation efforts. Ideally, the journal should have some degree of control over the archival platform linked to the proposals.

Thieberger et al. (2016) made dramatic progress in this line of discussion by systematically addressing the major “stumbling block” (so to quote Peter Austin’s comment to Haspelmath, Michaelis, 2014) to the realization of similar academic acknowledgement ideas, i.e., the lack of structured guidelines to review language data. Among the several points of their extensive proposal, Thieberger and colleagues stressed the importance of archiving raw (i.e., recordings) alongside primary (transcriptions) and structural (annotations) data while linking these layers and describing the documents through adequate sets of metadata. Following these works, the Linguistic Society of America (Fitzgerald, 2021) recently inaugurated a new section of its flagship *Language* journal, *Language Revitalization and Documentation*, dedicated to research ethics and methodology and, crucially, to corpus overviews. This effort was followed by a paper by Garrett and Harris (2022) on the assessment of scholarship in language documentation, in which archival collections are considered de facto research outputs. In the same years, but from the viewpoint of another linguistics sub-discipline, i.e., phonetics, the position paper by Garellek, Gordon, Kirby, Lee, Michaud, Mooshammer, Niebuhr, Recasens, Roettger, Simpson & Yu (2020) raised some perplexities on this emerging “dual reward system”. The authors argue that the separation of the scholarly acknowledgement of data curation from the value of research publications making use of the curated data risks engendering new weighting asymmetries in assessment systems, with archival publications being inevitably valued less and thus implicitly discouraged. Moreover, given the different research traditions of language documentation and phonetics, the authors fear that “giving additional credit to researchers for the archiving of phonetic data sets would feel like mixing apples with oranges” (ibid.: 11). On the contrary, in their view, data archiving should be recognized as a prerequisite for the submission of any phonetics research article to any journal.

While Garellek et al.’s (2020) point on valuing discrepancies does imply a constant monitoring of researcher (and institution) mindsets in assessing data contributions, their concern on diverging academic traditions does not match the current position held by the Italian research initiative behind the *Vademecum* (i.e., the *Tavolo permanente per le fonti orali*), which tries to raise awareness on crucial issues concerning human voice data through a work rooted in interdisciplinarity.

3.2.2 OAr and its structure

Against this background, the above-synthesized academic discussion on data journals was first brought to the attention of the *Vademecum* work group on data conservation in an in-person meeting held in Padua on July 10th, 2019. On that occasion, the idea of an Italian instantiation of this trend was deemed premature. Now the time seems ripe to work towards a new Italian journal open to oral archive data papers. Indeed, the state-of-the-art on data journal made further progress through the intervention of the Linguistic Society of America. Meanwhile, in Italy, the *Vademecum* has been made available to the community, providing valuable guidelines for oral data curation. Moreover, the *Archivio Vi.Vo.* platform at CLARIN-IT is entering a new phase of development which, through multiple tangential projects, will lead to its public release⁵. By offering a validated framework for oral document archiving, *Archivio Vi.Vo.* could help both contributors with their submissions and reviewers with their evaluations.

The University of Siena has recently sealed a deal with a National academic publishing house to host the annual issues of *Oral Archives Journal* (OAr). OAr will be structured in two main sections, for a total of eight subsections. Inspired by the original work groups of the *Vademecum*, the first main section is named *The life cycle of the archives* and welcomes contributions on

- a. Oral data production, encompassing topics such as conducting fieldwork, constructing an oral archive, and applying existing metadata schemes or developing new ones;
- b. Oral data curation, that is, data papers on oral archives deposited in reviewable repositories;
- c. Oral data conservation, dealing with strategies for the restoration of physical carriers and the speech signal, pipelines for digitization and data migration, and storage sustainability;
- d. Oral data use and re-use, hosting both original research papers and recounts of specific initiatives (e.g., exhibitions, touristic sound walks, etc.) making use of oral archives.

The second main section is focused on *Cross-cutting dimensions*, i.e., topics that are relevant to more than just one life phase of an oral archive:

- a. Speech transcription, from both theoretical and operative perspectives;
- b. Speech technology, including presentations of new tools for speech recording, processing, and archiving, as well as benchmarks and usability tests;

⁵ Currently, *Archivio Vi.Vo.* is in closed beta. Up to now, it has been used to process the archive of the Tuscan folk singer Caterina Bueno (Calamai et al., 2022) and the Anna Buonomini archive, mostly consisting of ethnological interviews conducted in the mountains of the *Provincia di Pistoia* (this was done during the *Grafo Reloaded* project; see Pozzebon, Biliotti & Calamai, 2016 for a brief, albeit slightly outdated, description of this archive). See Luzietti (2023) for a first evaluation of *Archivio Vi.Vo.* in its current state, which was performed through a processing test of the historian Angela Spinelli's oral archive.

- c. Legal issues, from consent procedures, privacy, data sensitivity, licensing, etc., up to copyright issues;
- d. Research ethics, fostering good practices for forging virtuous relationships between the researchers and the communities, including strategies for results dissemination and restitutionary work to the communities.

Given its peculiar relevance to the academic discussion leading to OAr (§ 3.2.1) and, more in general, to the ecosystem of oral archives, Oral data curation deserves additional explanation. Specifically, proposals submitted to this section will be evaluated following an adaptation of the rubric of Fitzgerald (2021: 5, based on Thieberger et al., 2016), reproduced below:

- a. Accessibility (deposited in a repository committed to providing long-term curation and access, including a persistent identifier and a citation form for items within the deposit; has a landing page or file with a basic description; includes access to metadata and a clear path to accessing the data in the corpus; files are in formats that are nonproprietary; levels of accessibility are properly justified from a legal standpoint).
- b. Quality (the nature and amount of contextual and background information; the structure of the deposit; metadata quality; the nature of linguistic annotation of the data; structural linking between raw data and their annotations, such as time-aligned transcriptions).
- c. Quantity (content; amount of data).

Compared to Fitzgerald's (2021) version, we revised the last point of the Accessibility section in order to stress that OAr does not endorse open access solutions which are not backed by a thorough legal feasibility study and, at the same time, does not necessarily reject archives with restricted access if this decision is properly motivated in their specific contexts of production. Moreover, OAr does not provide clear-cut interpretative guidelines to the Quantity section: the contribution of content and amount of data to the significance of the presented archive should be evaluated on a case-by-case basis.

3.2.3 Some alternatives and the specificity of OAr

The attentive reader may argue that, in addition to the above-mentioned Australian and American traditions, other lines of research in linguistics came to the establishment of renowned journals accepting descriptions of oral (but not only) data collections (linked to the actual data) as paper proposals. A notable example of this is *Language Resources and Evaluation* (Ide, Calzolari, 2005), which, under the auspices of the European Language Resources Association, is concerned with "language data and descriptions in machine readable form used to assist and augment language processing applications, such as written or spoken corpora and lexica, multimodal resources, [...], multimedia databases, etc. [...]"⁶.

Other outlets with broader scopes but interested in linguistics contributions are, for example, the *Journal of Open Humanities Data*, which welcomes short data

⁶ <https://link.springer.com/journal/10579>. Accessed 08 October 2024.

papers on “research object[s] with high reuse potential”, while stating that this type of work “does not replace a traditional research article, but rather complements it”⁷; and the *Research Data Journal for the Humanities and Social Sciences*, which is currently funded by the Consortium of European Social Science Data Archives and “aims to contribute to the transparency of research, accelerate dissemination and foster the reuse of scholarly data”⁸.

Descriptions of oral (or audiovisual) data can be read in journals conceived to appeal also to researchers from disciplines well beyond the humanities. *Data in Brief* hosts “short, digestible data articles”, “contributes to open science”, “improves reproducibility”⁹, and also stresses that “without a thorough description, any initial value to the data is lost” (Shaklee, 2014: 5). While phonetics was not originally mentioned among the fields exemplifying its “broad spectrum of disciplines” (that is, “including Biology, Chemistry, Economics, Psychology and Physics”; Wang, 2014: 85), articles describing oral data of related interest are evidently considered for publication (e.g., Di Benedetto, De Nardis, 2022). The similarly named journal *Data* have a “Data descriptors” section which serves analogous purposes; it should be noted that its “Data in application” subject area includes research and experimental data along with data in healthcare, finance, etc.¹⁰. Accounts of speech recorded for disparate ends and pertaining to heterogeneous disciplines can be found, e.g., in Ezzes, Schneck, Casilio, Fromm, Mefferd, de Riesthal & Wilson (2022) and Doyle, Şerban (2024).

This non-exhaustive list highlights a scientific trend kin to the intents of OAr while, on the other hand, reinforcing its specificity. In particular, OAr is centered around the orality of the discussed research data. This implies that the journal is not specifically tied to any individual discipline, and strives to be of interest to all the researchers involved with human voice (e.g., anthropologists, archivists, ethnomusicologists, jurists, linguists, oral historians, psychologists, sociologists, etc.). At the same time, this general theme excludes any description of data which is not oral (e.g., written or visual). OAr follows a line of archival discussion trying to go beyond the idea of oral data as cumbersome substitutes of transcriptions or limited to means of interpretation of written data (Valentini, Piccardi, Calamai & Stamuli, 2023). OAr not only puts oral data in the forefront, but also values them as documents before resources, i.e., the journal’s take on the documentary value of oral data does not necessarily end with any anticipation of their possible utilization in, e.g., technological or scientific endeavors¹¹. In accordance with the *Vademecum*, OAr primarily seeks to endorse an enhancement of the informative value of oral data.

⁷ <https://openhumanitiesdata.metajnl.com/about>. Accessed 08 October 2024.

⁸ <https://brill.com/view/journals/rdj/rdj-overview.xml>. Accessed 08 October 2024.

⁹ <https://www.sciencedirect.com/journal/data-in-brief/about/aims-and-scope>. Accessed 08 October 2024.

¹⁰ <https://www.mdpi.com/journal/data/about>. Accessed 08 October 2024.

¹¹ Note that this is not an understatement of Reusability as a core component of the FAIR principles: from OAr’s perspective, oral data have to be set up for their reuse (through, e.g., adequate metadata) as a prerequisite of a deserving archival process.

This implies that the archives submitted to OAr should present a well-thought-out conceptualization of the communicative events, exhaustive descriptions, and well-constructed networks of relationships between the oral documents and contextual information (such as related documents of other kinds – for example, a concert flyer or a picture of an interviewee – if any). In the case of documents from the pre-digital era, OAr favors effective linkages between the physical carriers of the data, their digital versions, and the individual communicative events. Lastly, OAr considers archive data papers as scholarly contributions on par with other types of articles (see § 3.2.1).

4. Closing remarks. Now it really comes down to trust

A dedicated journal represents a missing piece of the emerging picture of Italian tools promoting a paradigm shift towards data curation. Hopefully, knowledge and incentives will lower the perceived minimum requirements to set in motion the trusting virtuous circle of open science. Nonetheless, countering a major societal trend as the one described in § 1 may require other widespread actions. For this reason, we do hope that an impactful discussion on oral data management will germinate in Italian universities. Through the development of, e.g., pertinent course modules and handbooks, the students of today should be able to grow in a (research) world without any fear of FAIRness.

Acknowledgements

The *Gra.fo Reloaded* project was funded by Regione Toscana (Fondo per lo Sviluppo e la Coesione – FCS 2014-2020, Giovanisi), Siena University, Fondazione Sistema Toscana, and Ecomuseo della Montagna Pistoiese. Soprintendenza Archivistica e Bibliografica della Toscana, Istituto di Linguistica Computazionale “Antonio Zampolli” (Pisa National Research Council) and CLARIN-IT also contributed to the project. The authors would like to thank all who have taken the Tuscan census into consideration and Maria Francesca Stamuli (Soprintendenza Archivistica e Bibliografica della Toscana) for having brought Mulè’s (1993) passage on nonresponse to their attention.

References

- ANDREINI, A., CLEMENTE, P. (Eds.) (2007). *I custodi delle voci. Archivi orali in Toscana: primo censimento*. Florence: Regione Toscana.
- BAMBERGER, M.R., REEVES, T.D. (2021). Individual and contextual factors associated with data sharing in the social sciences. In *The Social Science Journal*. Advance online publication. <https://doi.org/10.1080/03623319.2021.1986663>.

- BAUER, P.C., FREITAG, M. (2018). Measuring trust. In USLANER, E.M. (Ed.), *The Oxford Handbook of Social and Political Trust*. New York: Oxford University Press, 15-36. <https://doi.org/10.1093/oxfordhb/9780190274801.013.1>.
- BELLIARD, F., MAINERI, A., PLOMP, E., RAMOS PADILLA, A.F., SUN, J. & JEDDI, M.Z. (2023). A 10 step checklist for starting FAIR discussions in your community: Call for contributions. In *Fair Connect*, 1(1), 45-48. <https://doi.org/10.3233/FC-230505>.
- BENSON-GREENWALD, T.M., TRUJILLO, A., WHITE, A.D. & DIEKMAN, A.B. (2023). Science for others or the self? Presumed motives for science shape public trust in science. In *Personality and Social Psychology Bulletin*, 49(3), 344-360. <https://doi.org/10.1177/01461672211064456>.
- BOCHYNSKA, A., KEEBLE, L., HALFACRE, C., CASILLAS, J.V., CHAMPAGNE, I-A., CHEN, K., RÖTHLISBERGER, M., BUCHANAN, E.M. & ROETTGER, T.B. (2023). Reproducible research practices and transparency across linguistics. In *Glossa Psycholinguistics*, 2(1), 1-36. <https://doi.org/10.5070/G6011239>.
- BREHM, J., SABEL, M. (2019). What do survey measures of trust actually measure?. In SASAKI, M. (Ed.), *Trust in Contemporary Society*. Leiden/Boston: Brill, 233-260. https://doi.org/10.1163/9789004390430_013.
- BRUDNER, E.G., KAROUSATOS, A.J., FARERI, D.S. & DELGADO, M.R. (2022). Trust and reputation. How knowledge about others shapes our decisions. In KRUEGER, F. (Ed.), *The Neurobiology of Trust*. Cambridge: Cambridge University Press, 155-184. <https://doi.org/10.1017/9781108770880.010>.
- CALAMAI, S., CASELLATO, A. & STAMULI, M.F. (2022). Collaborative best practices. An Italian Vademecum on the conservation, the description, the use and the re-use of oral sources. In *Bulletin de l'AFAS. Sonorités*, 48, 182-195. <https://doi.org/10.4000/afas.7499>.
- CALAMAI, S., PICCARDI, D., PRETTO, N., CANDEO, G., STAMULI, M.F. & MONACHINI, M. (2022). Not just paper: Enhancement of archive cultural heritage. In FIŠER, D., WITT, A. (Eds.), *CLARIN. The Infrastructure for Language Resources*. Berlin/Boston: De Gruyter, 647-665. <https://doi.org/10.1515/9783110767377-025>.
- CITRIN, J., STOKER, L. (2018). Political trust in a cynical age. In *Annual Review of Political Science*, 21, 49-70. <https://doi.org/10.1146/annurev-polisci-050316-092550>.
- CLEMENTE, M.A. (2023). Genova ferita, una città all'interno di un evento internazionale. Master of Arts dissertation, Università di Pisa.
- CONTI, N., MEMOLI, V. (2015). Show the money first! Recent public attitudes towards the EU in Italy. In *Italian Political Science Review*, 45(2), 203-222. <https://doi.org/10.1017/ipo.2015.11>.
- DE JONG, F., MAEGAARD, B., FIŠER, D., VAN UYTVANCK, D. & WITT, A. (2020). Interoperability in an infrastructure enabling multidisciplinary research: the case of CLARIN. In CALZOLARI, N., BÉCHET, F., BLACHE, P., CHOUKRI, K., CIERI, C., DECLERCK, T., GOGGI, S., ISAHARA, H., MAEGAARD, B., MARIANI, J., MAZO, H., MORENO, A., ODIJK, J. & PIPERIDIS, S. (Eds.), *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*. Paris: ELRA, 3406-3413.
- DI BENEDETTO, M-G., DE NARDIS, L. (2022). The GEMMA speech database: VCV and VCCV words for the acoustic analysis of consonants and lexical gemination in Italian. In *Data in Brief*, 43. <https://doi.org/10.1016/j.dib.2022.108373>.

- DIJKERS, M.P. (2019). A beginner's guide to data stewardship and data sharing. In *Spinal Cord*, 57, 169-182. <https://doi.org/10.1038/s41393-018-0232-6>.
- DOYLE, D., ȘERBAN, O. (2024). Interruption audio & transcript: derived from group affect and performance dataset. In *Data*, 9(9). <https://doi.org/10.3390/data9090104>.
- EPHRATT, M. (2008). The functions of silence. In *Journal of Pragmatics*, 40(11), 1909-1938. <https://doi.org/10.1016/j.pragma.2008.03.009>.
- EZZES, Z., SCHNECK, S.M., CASILIO, M., FROMM, D., MEFFERD, A.S., DE RIESTHAL, M. & WILSON, S.M. (2022). An open dataset of connected speech in aphasia with consensus ratings of auditory-perceptual features. In *Data*, 7(11). <https://doi.org/10.3390/data7110148>.
- FERRITER, D. (1998). Oral archives in Ireland: a preliminary report. In *Irish Economic and Social History*, 25(1), 91-95. <https://doi.org/10.1177/033248939802500108>.
- FICHTNER, U.A., HORSTMAYER, L.M., BRÜHMANN, B.A., WATTER, M., BINDER, H. & KNAUS, J. (2023). The role of data sharing in survey dropout: a study among scientists as respondents. In *Journal of Documentation*, 79(4), 864-879. <https://doi.org/10.1108/JD-06-2022-0135>.
- FITZGERALD, C.M. (2021). A framework for Language Revitalization and Documentation. In *Language*, 97(1), 1-11. <https://doi.org/10.1353/lan.2021.0006>.
- GARELLEK, M., GORDON, M., KIRBY, J., LEE, W-S., MICHAUD, A., MOOSHAMMER, C., NIEBUHR, O., RECASENS, D., ROETTGER, T.B., SIMPSON, A. & YU, K.M. (2020). Letter to the editor: Towards open data policies in phonetics: What we can gain and how we can avoid pitfalls. In *Journal of Speech Sciences*, 9, 3-16. <https://doi.org/10.20396/joss.v9i00.14955>.
- GARRETT, A., HARRIS, A.C. (2022). Assessing scholarship in documentary linguistics. In *Language*, 98(3), 156-172. <https://doi.org/10.1353/lan.0.0266>.
- GOMES, D.G.E., POTTIER, P., CRYSTAL-ORNELAS, R., HUDGINS, E.J., FEROUGHIRAD, V., SÁNCHEZ-REYES, L.L., TURBA, R., MARTINEZ, P.A., MOREAU, D., BERTRAM, M.G., SMOUT, C.A. & GAYNOR, K.M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. In *Proceedings of the Royal Society B*, 289(1987). <https://doi.org/10.1098/rspb.2022.1113>.
- HARDIN, R. (2006). *Trust*. Cambridge/Malden: Polity Press.
- HARSIN, J. (2018). Post-truth and critical communication studies. In *Oxford Research Encyclopedia of Communication*. Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780190228613.013.757>.
- HASPELMATH, M., MICHAELIS, S.M. (2014). Annotated corpora of small languages as refereed publications: a vision. *Diversity Linguistics Comment*. <https://dlc.hypotheses.org/691>. Accessed 08 October 2024.
- IDE, N., CALZOLARI, N. (2005). Introduction to the special inaugural issue. In *Language Resources and Evaluation*, 39(1), 1-7. <https://doi.org/10.1007/s10579-005-2689-0>.
- JANDA, L.A. (2021). Managing data and statistical code according to the FAIR principles. In BEREZ-KROEKER, A.L., McDONNELL, B., KOLLER, E. & COLLISTER, L.B. (Eds.), *The Open Handbook of Linguistic Data Management*. Cambridge/London: The MIT Press, 447-452. <https://doi.org/10.7551/mitpress/12200.003.0042>.

- LAMBIN, J.-J. (2008). *Changing market relationships in the internet age*. Louvain-la-Neuve: Presses universitaires de Louvain.
- LAWRENCE, B., JONES, C., MATTHEWS, B., PEPLER, S. & CALLAGHAN, S. (2011). Citation and peer review of data: moving towards formal data publication. In *International Journal of Digital Curation*, 6(2), 4-37. <https://doi.org/10.2218/ijdc.v6i2.205>.
- LUZIETTI, R.B. (2023). Evaluation of the Archivio Vi.Vo architecture: A case study on the reuse of legacy data for linguistic purposes. In ERJAVEC, T., ESKEVICH, M. (Eds.), *Selected Papers from the CLARIN Annual Conference 2022*. Linköping: Linköping University Electronic Press, 90-98. <https://doi.org/10.3384/ecp198009>.
- MANALI, S. (2024). Il ritorno della Pantera. Un laboratorio universitario per la costruzione di un archivio orale. In *JLIS*, 15(1), 159-176. <https://doi.org/10.36253/jlis.it-571>.
- MULÈ, A. (1993). Censimento e presentazione dei dati. In BARRERA, G., MARTINI, A. & MULÈ, A. (Eds.), *Fonti orali. Censimento degli istituti di conservazione*. Roma: Ministero per i beni culturali e ambientali, Ufficio centrale per i beni archivistici, 33-49.
- PICCARDI, D., ARDOLINO, F. & CALAMAI, S. (Eds.) (2019). *Audio archives at the crossroads of speech sciences, digital humanities and digital heritage*. Studi AISV 6. Milano: Officinaventuno.
- PICCARDI, D., CALAMAI, S. (2023). How many oral archives are in your home? Piloting a new Tuscan census in the Gra.fo Reloaded project. In CARBÉ, E., LO PICCOLO, G., VALENTI, A. & STELLA, F. (Eds.), *La memoria digitale. Forme del testo e organizzazione della conoscenza. Atti del XII convegno annuale AIUCD*. University of Siena: AIUCD, 389-394.
- POZZEBON, A., BILIOTTI, F. & CALAMAI, S. (2016). Places speaking with their own voices. A case study from the Gra.fo archives. In IOANNIDES, M., FINK, E., MOROPOULOU, A., HAGEDORN-SAUPE, M., FRESA, A., LIESTØL, G., RAJCIC, V. & GRUSSENMEYER, P. (Eds.), *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. EuroMed 2016*. Cham: Springer, 232-239.
- ROTENBERG, K.J. (2018). *The psychology of trust*. London and New York: Routledge. <https://doi.org/10.4324/9781315558912>.
- SHAKLEE, P.M. (2014). Data in brief – Making your data count. In *Data in Brief*, 1, 5-6. <https://doi.org/10.1016/j.dib.2014.09.001>.
- STRACKE, C.M. (2020). Open science and radical solutions for diversity, equity and quality in research: a literature review of different research schools, philosophies and frameworks and their potential impact on science and education. In BURGOS, D. (Ed.), *Radical Solutions and Open Science*. Singapore: Springer, 17-37. https://doi.org/10.1007/978-981-15-4276-3_2.
- TAVOLO PERMANENTE PER LE FONTI ORALI (2023). *Vademecum per il trattamento delle fonti orali*. Roma: Ministero della Cultura, Direzione generale Archivi.
- THIEBERGER, N. (2012). Counting collections. Endangered Languages and Cultures. <https://www.paradisec.org.au/blog/2012/11/counting-collections/>. Accessed 08 October 2024.
- THIEBERGER, N., MARGETTS, A., MOREY, S. & MUSGRAVE, S. (2016). Assessing annotated corpora as research output. In *Australian Journal of Linguistics*, 36(1), 1-21. <https://doi.org/10.1080/07268602.2016.1109428>.
- VALENTINI, C., PICCARDI, D., CALAMAI, S. & STAMULI, M.F. (2023). Da cornice a soggetto. Il documento sonoro nell'infrastruttura Archivio Vi.Vo. In *Archivi*, 18(1), 88-125.

WANG, H-R. (2014). "Publish or perish": Should this still be true for your data? In *Data in Brief*, 1, 85-86. <https://doi.org/10.1016/j.dib.2014.11.005>.

WILKINSON, M.D., DUMONTIER, M., AALBERSBERG, I.J., APPLETON, G., AXTON, M., BAAK, A., BLOMBERG, N., BOITEN, J- W., DA SILVA SANTOS, L.B., BOURNE, P.E., BOUWMAN, J., BROOKES, A.J., CLARK, T., CROSAS, M., DILLO, I., DUMON, O., EDMUNDS, S., EVELO, C.T., FINKERS, R., ... MONS, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. In *Scientific Data*, 3. <https://doi.org/10.1038/sdata.2016.18>.

