Nomination for the 2024 Dutch Data Prize (SSH domain)

1. Your organisation
Humanities Cluster, Royal Netherlands Academy of Arts and Sciences (KNAW)

2. Your name and surname
Prof.dr. Inger Leemans (PI of the EU Horizon 2020 *Odeuropa* project)

3. Your e-mail
Inger.Leemans@huc.knaw.nl

4. Name data(set)
European Olfactory Knowledge Graph (EOKG)

5. Persistent Identifier / URL
https://doi.org/10.5281/zenodo.10709703 (dataset, code)
https://github.com/Odeuropa/knowledge-graph/ (dataset, code)

The European Olfactory Knowledge Graph can be explored via:

- **EOKG Vocabularies**: https://vocab.odeuropa.eu/ (vocabulary browser)
- **Odeuropa Ontology**: https://data.odeuropa.eu/ontology/ (data model)
- **EOKG API**: https://grlc.eurecom.fr/api/Odeuropa/kg-api/ (API)
- **EOKG technical report:** https://doi.org/10.3030/101004469 (documentation)
- **Odeuropa Smell Explorer:** https://explorer.odeuropa.eu/ (demonstrator)

See also the *Odeuropa* project website: https://odeuropa.eu

6. Description of the data(set) in up to 100 words
The European Olfactory Knowledge Graph (EOKG) is an innovative gateway into the history and heritage of smell and olfactory practices in Europe, consisting of data, data models, and vocabularies related to the role of smell in culture. The data was extracted from a large set of European cultural heritage collections: 43,000 images and 167,000 texts in six languages (English, Dutch, Italian, French, German and Slovene) from the 17th to 20th century. It provides over 2.5M references to historical smell experiences, including:

- olfactory objects ('things that smell')
- olfactory spaces
- olfactory gestures
- olfactory iconographies
- noses (expert smellers)
- emotions evoked by smell

7. Name of the contact person for the data(set)
Raphaël Troncy, EURECOM, France (https://www.eurecom.fr/en/people/troncy-raphael)

8. E-mail of the contact person of the dataset
raphael.troncy@eurecom.fr

9. The data(set) qualifies for the category
Social Sciences and Humanities

10. Are the data widely used or have the potential to be so?
Yes

11. Please elaborate

The Odeuropa project has created several different gateways into the EOKG which can be accessed from the 'Results' section of its website (https://odeuropa.eu).

The most important of these is the *Odeuropa Smell Explorer* demonstrator, a combined search and browse interface and our main demonstrator for the EOKG data and ontology. In the last 12 months, it has been accessed by over 7300 unique users. The *Encyclopedia of Smell History and Heritage* makes use of and links back to the EOKG. Several interactive Jupyter 'nosebooks' (hosted on Google Collab) show researchers how to query the data, and we also developed a serious game (*Nefertum*) which allows one to play around with the olfactory vocabularies.

The EOKG data and data model have also been disseminated through many publications (https://odeuropa.eu/publications/) including an award winning 2022 paper, "Capturing the Semantics of Smell: The Odeuropa Data Model for Olfactory Heritage Information", and an extensive series of workshops, presentations and other public events, most notably the *Odeuropa Smell Culture Fair* (https://odeuropa.eu/smell-culture-fair-programme/) at the KNAW Trippenhuis in November, 2023, which attracted over 130 participants. We created a teaching module, "Knowing by Sensing: How to Teach the History of Smell" for the *American Historical Review* that drew on the data in the EOKG. It consists of a series of videos and documents that explain how the history of smell can be taught in a 'nose-on' fashion to higher-education students. Over the last three years, we have taught the project's methods and shared our data at 15 courses to circa 760 students and PhD candidates.

Although the Odeuropa project just ended in December, 2023 we can already show some concrete examples of scientific exploitation of its results by other researchers. In the Netherlands, *POEM*, a 2023 NWO Open Competition funded project led by dr. J.F. van Dijkhuizen (Leiden University) is drawing on the EOKG to help it analyse the literary function of smell perception in early modern poetry and drama. In June this year, the University of Geneva in Switzerland organised a Hackathon where participants built a demo application showing on a map the convergence of olfactory experiences and emotions, highlighting the valence of the emotions themselves in geographic areas. The goal of this demo was to compare more conceptual olfactory spaces (e.g. urban/rural, indoor/outdoor) with olfactory emotional perceptions in specific geographic regions in Europe and the rest of the world.

Crucially, we do not consider our work on Odeuropa to be 'done'. In November this year, several members of the consortium will apply for an EU MSCA Doctoral Network grant to train new Ph.D. students in smell heritage and olfactory data analysis. In addition, the Odeuropa team is developing an olfactory consultancy group whose goal will be to support and advise cultural heritage institutes and other clients in the GLAM sector in preparing olfactory exhibits and other forms of olfactory storytelling. All of these initiatives will bank on the data and knowledge of the EOKG and Smell Explorer.

12.Are the data openly available, or does the dataset documentation explain why this is not possible and under which conditions interested parties can get access to the data?

All the code and data are freely available under open access and open source licenses.

13.Please elaborate

The hosted vocabularies and ontologies can be searched and explored via https://vocab.odeuropa.eu/en/ and https://data.odeuropa.eu. The underlying datasets are available via the previously linked Github repositories and preserved on the EU-funded Zenodo platform, https://zenodo.org/communities/odeuropa/ The EOKG has a CC-BY license. The associated documentation and media created in the Odeuropa project is shared under a Creative Commons CC-BY 4.0 license.

Per EU data management rules, all project websites, data and online services (such as demonstrators) hosted by the Odeuropa partners are being maintained for at least five years past the end of the project (i.e. December 2028). Following this, if for any reason the data and services cannot be kept online by the project partners beyond this period, copies of the data and code will be permanently archived with a Core Trust Seal certified digital repository in the Netherlands (e.g. KNAW DANS Data Station for SSH or the KNAW IISG Dataverse). Further information can be found in the Odeuropa Data Management Plan: https://odeuropa.eu/wp-content/uploads/2022/05/D1_4_DataManagementPlan.pdf

14. Are the data or the research opportunities it brings of importance?
Please score the following:

4 - Important for the entire domain of research and beyond
3 - Important for closely related disciplines
2 - Important for research within the same discipline
1 - Unlikely to be important to others outside of the immediate research group

15.Please elaborate
The research and exploitation opportunities of the EOKG are manifold and extend to disciplines well beyond the social sciences and humanities such as cognitive psychology, computer science, and mechanical engineering. This is because the data and models of the EOKG can help bridge the gap between an analytic, quantitative characterisations of odours based on their chemistry, and the qualitative, linguistic description of, for example, odour-sentiment relations described in olfactory experiences. In fact, odour mapping is one of the biggest challenges for companies working in the field of aroma flavour research since it requires the creation and interpretation of structured datasets by specialists drawn from a wide range of disciplines (cultural history, linguistics, chemistry, cognitive sciences, food science, etc).

A further area for reuse and exploitation outside the SSH domain is in the odour and fragrance industry, and in the AI-driven recommendation services that are now increasingly being used to collect and analyse data around consumer preferences. For example, Odeuropa data can provide insights to industry researchers on long-term, historical, linguistically segmented fragrance preferences and the specific type of language and sentiment associated with them. This data, in turn, is then used to better understand and characterise what kinds of ‘scent stories’ should be marketed to potential customers. For examples, see the Odeuropa-related articles in industry publications such as Fragrantica, Essencional, Perfumer & Flavorist, Basenotes, and Nez (and, more broadly, https://odeuropa.eu/press/).

16.Does the dataset show strong awareness of items 4, 6, 7, 8, and 9 of https://fairaware.dans.knaw.nl?

Yes

17. Please elaborate

Relevant Fairware Criteria:

4. Are you aware that access to your data(set) may need to be controlled and that metadata should include licence information under which the data(set) can be reused?
Yes
6. Are you aware that the metadata describing your data(set) should use controlled vocabularies?
Yes
7. Are you aware that provenance information about the collection and/or generation of data should be included in the metadata?
Yes
8. Are you aware that metadata describing your data(set) should follow the specifications of a community-endorsed standard?
Yes
9. Are you aware that your data(set) should be deposited preferably in a file format that is open and supported by the data repository for long-term preservation?
Yes

18.Are the data in a digital repository that is certified as trustworthy (demonstrated through CoreTrustSeal (https://www.coretrustseal.org) or World Data System (https://www.worlddatasystem.org/services/certification) membership or in a public digital repository that shows strong awareness of items 1, 2, 3, 5, and 10 of https://fairaware.dans.knaw.nl ?

Yes

19. Please elaborate

Our data is already being preserved on Zenodo which complies with both the FAIR and the EU's Plan-S principles (https://about.zenodo.org/principles/). In addition, we plan to deposit a second, permanent archival copy of our data and code in Core Trust Seal certified digital repository in the Netherlands (e.g. the KNAW DANS Data Station for SSH or the KNAW IISG Dataverse). For more information on the Odeuropa Data Management Plan, please see: https://odeuropa.eu/wp-content/uploads/2022/05/D1_4_DataManagementPlan.pdf

Bibliography

Odeuropa Technical Deliverables

https://odeuropa.eu/deliverables/
https://doi.org/10.3030/101004469 (EU CORDIS database)

Odeuropa Publications

https://odeuropa.eu/publications/

Odeuropa Press Coverage

https://odeuropa.eu/press/