

HEREDITARY

HetERogeneous sEmantic Data integration for the guT-bRain interplaY

Deliverable 5.2

Demonstrator of Visualization Components for Sequences, Networks, Text, and High-Dimensional Data

Version 1.10, 23.12.2025

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.



**Funded by
the European Union**

EXECUTIVE SUMMARY

In HEREDITARY's Work Package 5 (WP5) we develop and evaluate data visualization and information extraction methods, to support analysis and understanding of multimodal biomedical data. In the Deliverables 5.1 and 5.3 of Month 12, based on initial requirement analysis with project partners, visualization components covering a wide range of data modalities were developed. In the course of HetERogeneous sEmantic Data integration for the guT-bRain interplaY (HEREDITARY), these components will gradually be developed into data visualization applications, for application and evaluation. As a key step in this direction, we develop an application server that hosts the visualization components, and which can demonstrate the components, and serve to apply the components to project data by partners. A key goal is to make our applications openly available, including source code. This deliverable describes our approach to open practice, the architecture of our application server, and the first set of applications we deployed.

DOCUMENT INFORMATION

Deliverable ID	5.2
Deliverable Title	Demonstrator of Visualization Components for Sequences, Networks, Text, and High-Dimensional Data
Work Package	WP 5
Lead Partner	TU Graz
Due Date	31.12.2025
Date of submission	23.12.2025
Deliverable Type	DEM – Demonstrator
Dissemination level	PU – Public

AUTHORS

Name	Organization
Tobias Schreck	TUGRAZ
Stefan Lengauer	TUGRAZ
Peter Waldert	TUGRAZ
Benedikt Kantz	TUGRAZ
Casper van Leeuwen	SURF
Lukas Schilcher	TUGRAZ
Michael Grabner	TUGRAZ
Giacomo Nebbia (internal reviewer)	UCD
Anna Romanovych (contributor)	UNIPD
Gianmaria Silvello (contributor)	UNIPD

REVISION HISTORY

Version	Date	Author	Description
0.10	2025-02-28	Peter Waldert	Structure for deliverable
0.20	2025-11-25	All authors	Complete draft including all components
0.30	2025-12-12	All authors	Revised draft after internal review
1.00	2025-12-17	All authors	Final version
1.10	2025-12-17	Gianmaria Silvello	Minor fixes, layout check and submitted version

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Contents

1	Introduction	8
2	Component FAIRification	8
2.1	Findable	9
2.2	Accessible	9
2.3	Interoperable	9
2.4	Reusability	9
3	Demonstrator System	10
4	Demonstrators	12
4.1	Clusters in Focus	12
4.2	Neurodegen-Vis	13
4.3	OnSET – Ontology and Semantic Exploration	14
4.4	Gut-Brain Viewer	15
4.5	AlloyInter	16
5	Demonstration Video Material	17
6	Conclusion and Next Steps	18
	References	19

List of Figures

1	The demonstrator deployment pipeline.	11
2	Clusters in Focus is an application for comparing groups i data, and select features of interest. We provide the <i>Clusters in Focus</i> GitHub repository (Schilcher et al., 2025b), which documents tool usage and setup.	12
3	Screenshot of the currently deployed neurodegen-vis application.	13
4	Screenshot of the currently deployed Ontology and Sparse data Exploration Tool (OnSET).	14
5	Screenshot of the currently deployed Gut-Brain Viewer.	15
6	Screenshot of the currently deployed AlloyInter tool.	16
7	Demonstration videos for main components developed in the project.	17

Acronyms

CI Continuous Integration. 11

DOI Document Object Identifier. 9

DP Differential Privacy. 13

FAIR Findable Accessible Interoperable Reusable. 8

HEREDITARY HetERogeneous sEmantic Data integratlon for the guT-bRain interplaY. 3, 8, 10, 17

KG Knowledge Graph. 14

LLM Large Language Model. 10, 17

ML Machine Learning. 10

MRI Magnetic Resonance Imaging. 17

OnSET Ontology and Sparse data Exploration Tool. 6, 9, 14

WP5 Work Package 5. 3

1 Introduction

In HEREDITARY WP5 we develop and evaluate data visualization and information extraction methods, to support analysis and understanding of multimodal biomedical data. In the deliverables 5.1 and 5.3 (Schreck, Lengauer, Waldert, Kantz, Grabner, Schilcher, Percic, & van Leeuwen, 2024; Schreck, Lengauer, Waldert, Kantz, Grabner, Schilcher, Percic, van Leeuwen, et al., 2024) of month 12, based on initial requirement analysis with project partners, visualization components covering a wide range of data modalities were developed. In the course of HEREDITARY, these components will gradually be developed into data visualization applications, for application and evaluation. As a key step in this direction, we develop an application server that hosts the visualization components, and which can demonstrate the components, and serve to apply the components to project data by partners. A key goal is to make our applications openly available, including source code. This deliverable describes our approach to open practice, the architecture of our application server, and the first set of applications we deployed.

First, in section 2, we describe our approach for our developments open access and following open science standards, governed by the so-called Findable Accessible Interoperable Reusable (FAIR) principles. In section 3 we detail the architecture of our application demonstration server. In section 4, we describe the first deployed applications, which are based on components developed in the project. These include cluster analysis, knowledge graph exploration, multimodal image based visualization, and simulation data exploration. To demonstrate components and applications, he also produced demonstration videos, described in section 5. Finally, in section 6 we discuss next steps.

2 Component FAIRification

Before we detail our deployment approach and deployed systems, we lay out our strategy to manage and preserve the code developed for the visualizations, and by extension, the other systems developed within HEREDITARY, complying with the FAIR principles. Adapting the FAIR principles to software requires considerations, which are outlined in depth by Lamprecht et al. (2019). This is required, as research *software* differs to research *data*. While both can be linked to a unique identifier, archived, and made available through licenses, they differ in the way they are reused. Software is usually bound to workflows, and depends (most of the time) on dependencies that might become unavailable or deprecated, whereas data can stay FAIR for much longer. The core consideration for software to be FAIR lies therefore in the prolonging of use within a community (Lamprecht et al., 2019).

At the heart of our source management strategy are three goals, which align well with the FAIR principles of the project to publish our research in open-access formats:

1. We provide access to the code behind our (visualization) systems in an open and central manner.
2. The source for these systems stays accessible and reusable through permissible licenses beyond the project lifetime.
3. The code is reusable and citable in academic settings.

We will address these three goals through a set of recommendations for source management within the project, which we are now beginning FAIR implement in our deployed systems. These are laid out within the FAIR goals next.

2.1 Findable

Metadata We foster the findability of our demonstrator software components by utilizing Zenodo (Peters et al., 2017) to publish the metadata of our software components in a unified manner, also assigning a Document Object Identifier (DOI) to them allowing further downstream users to reference them easily. This implicitly addresses our third goal while attaching rich and readable metadata to the software in accordance with the FAIR goals (Lamprecht et al., 2019).

2.2 Accessible

Source Management Our first goals can be further achieved through an effort initiated within the project at Aalborg University. They have created a public GitHub organization¹ open to anyone. Project members can join and publish code under a centralized domain using this organization, and third parties can view and reuse the code. This should make our code both *findable* and *accessible* (Lamprecht et al., 2019).

2.3 Interoperable

We design our systems from the ground up to support open standards for data ingestion, such as RDF (FAIR-sharing Team, 2024) within OnSET (Kantz, Innerebner, et al., 2025a, 2025b) or CSV (FAIRsharing Team & McQuilton, 2024) within Neurodegen-Vis or Clusters In Focus (Schilcher et al., 2025a). This comes from the necessity to apply these systems on multiple modalities and datasets, therefore requiring to be open by default.

2.4 Reusability

Licensing of our Systems The next aspect of the license choice contributes to our second goal: improving downstream reusability for other users and applications, as well as possible extensions by third parties. We enable these efforts by attaching a license to the repositories in the LICENSE file. We have started using the MIT license (Laurent, 2004) because it is permissive and open, aligning with our goals of allowing almost any downstream use and satisfying the *reusability* requirements from a licensing perspective.

Documentation Furthermore, good documentation also improves downstream reusability for third parties, enabling them to reproduce these results easily. To this end, we provide instructions for others to get the software started.

Containerization Finally, we also provide instructions for containerizing applications for easy deployment within our systems (see the next Section), and for third parties to test these systems in their environments, even within specific sandboxes to guarantee data privacy.

¹ <https://github.com/hereditary-eu>

3 Demonstrator System

Our approaches and goals outlined in section 2 enable us to easily deploy our applications for public access. We use a specifically set up server system (we call *Demo System*), operated at TU Graz, to host our visualization components. This deployment strategy can be utilized across all of our components, as we use standard web services. We, nevertheless, enable a flexible technology stack, as containerization allows us to deploy components on various frameworks and web standards.

The deployment pipeline is shown in fig. 1. Our approach begins by mirroring the source repository from the HEREDITARY GitHub group into our internal *GitLab* group to manage deployments internally. These mirrored source repositories trigger a new pipeline run upon a new commit. This, in turn, builds a new Docker container to deploy the component to any location. The resulting containers are stored in the GitLab container repository, enabling easy deployment in the following steps or to arbitrary other systems.

These containers can then be pulled by the *Demo System* for the final deployment. The *Demo System* is configured with reliability and Machine Learning (ML) in mind, catering to (ideally) all of our demonstrator component needs within the HEREDITARY WP 5. We specifically configured it with

- a 12-core CPU,
- 64 GB ECC memory,
- 2 NVIDIA RTX 5070 Ti GPUs with 16 GB memory each.

This configuration enables us to deploy up to two 8B Large Language Models (LLMs) with Q8 quantization², while being cost-effective. These GPUs are already used for the systems in ?? and section 4.3, using a shared vLLM instance. We use the NVIDIA container runtime to manage allocation for the containers requiring GPU resources for ML inference.

Finally, the *Demo System* uses *traefik* to manage the ingress traffic from the Web. We configure multiple subdomains to point to the system. *traefik* additionally routes traffic to individual containers, enabling us to configure arbitrary subdomains to match them. We also manage our SSL certificates using *traefik* to provide encrypted connections.

The containers produced in the earlier pipeline stages are finally deployed using Docker compose files for easy reproducibility of the setup. Each compose file defines the required containers, possible GPU requirements, network, and disk bindings.

²Assuming 1.5 bytes per parameter on Q8 quantization, resulting in 12 GB usage with additional space for query caching directly on the GPU

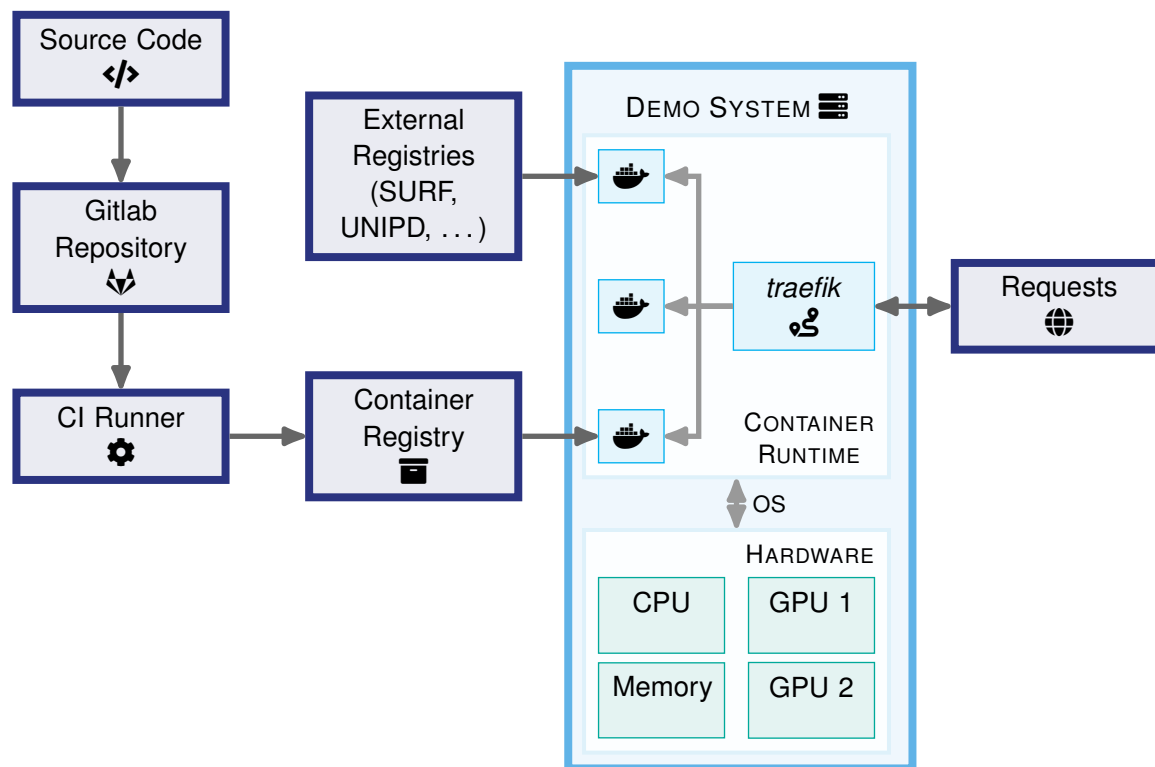


Figure 1: The demonstrator deployment pipeline at TU Graz. We use a Continuous Integration (CI) approach to build our source, and use containers to deploy it on our demonstrator system. We can also host arbitrary containers from partners as long as their repositories are accessible. The ingress is provided and managed using *traefik*.

4 Demonstrators

We describe the applications that are available at the demonstration system at this time.

4.1 Clusters in Focus

Clusters in Focus is a visualisation tool for high-dimensional patient data that enables the re-identification of (abstract) subgroups within the data. *Clusters in Focus* has been presented at the Eurographics Symposium on Visual Computing in Biology and Medicine (VCBM) 2025 in Delft (Schilcher et al., 2025a) and its code is made available publicly using the publication flow from above (Schilcher & Waldert, 2025). The README file contains instructions for setting up the tool locally or also in production deployment settings (based on *Docker*). A preview image can be found in Figure 2.

The tool is deployed in its current state on the *Demo System* and can be accessed via <https://hereditary.cgv.tugraz.at/clusters-in-focus>.

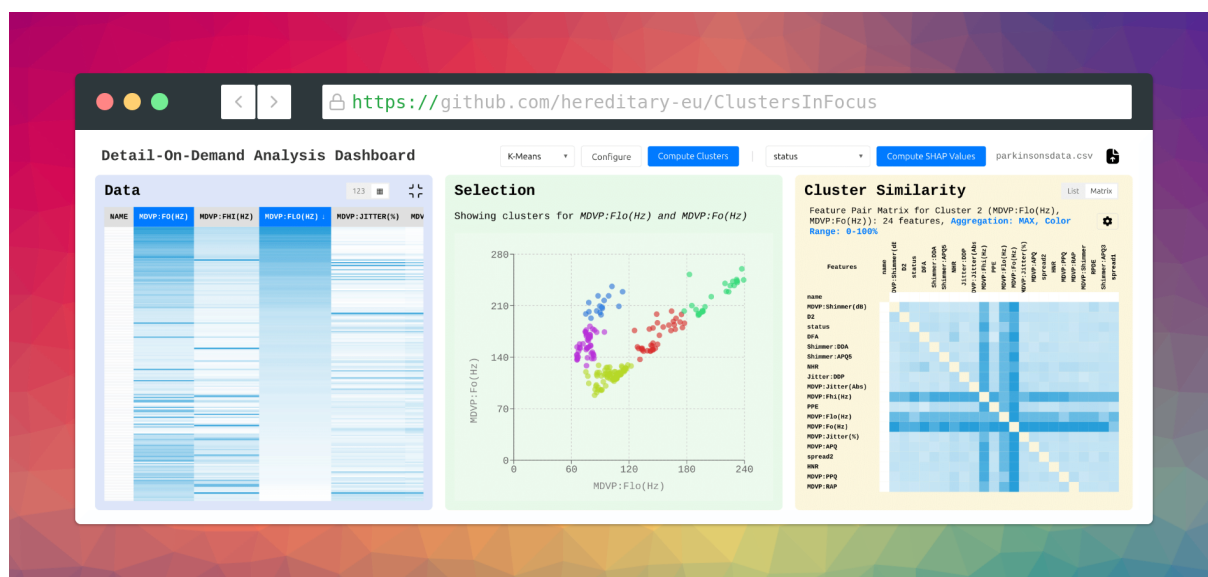


Figure 2: Clusters in Focus is an application for comparing groups i data, and select features of interest. We provide the *Clusters in Focus* GitHub repository (Schilcher et al., 2025b), which documents tool usage and setup.

4.3 OnSET – Ontology and Semantic Exploration

OnSET (Kantz, Innerebner, et al., 2025a, 2025b) is also deployed at the *Demo System* in its most current state. OnSET is a query tool providing a natural language interface for Knowledge Graphs (KGs) querying, integrated with a visual node-link builder for query adjustments. The system's core contribution is the addition of semantic tooling into ontology exploration and query construction informed by ontologies. Further details can be found in deliverable 5.5 (Schreck et al., 2025). We have indexed the code using Zenodo (Kantz, 2025b). The system requires at least five containers³, which are easily managed using the compose file. The system is available on <https://onset.cgv.tugraz.at/>, with a screenshot shown in fig. 4.

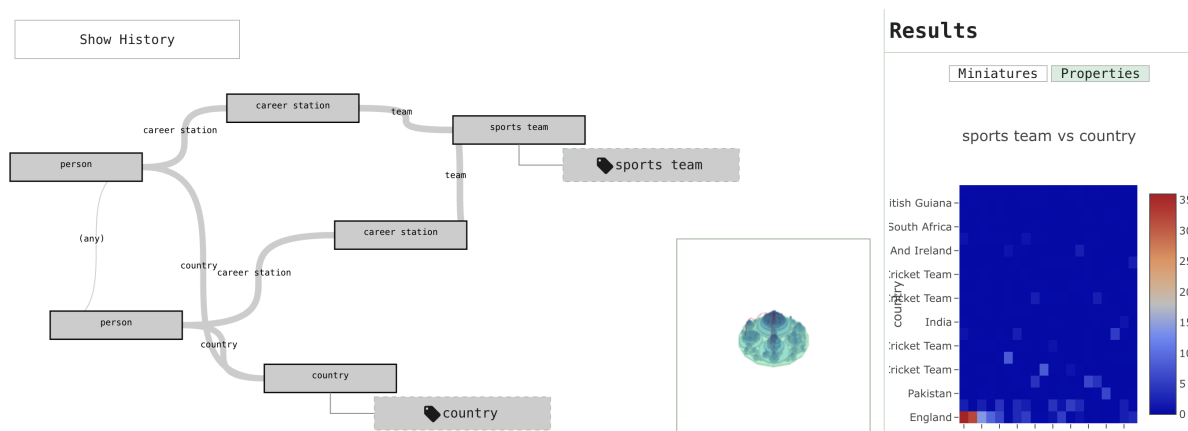


Figure 4: Screenshot of the currently deployed OnSET.

³The frontend, backend, Postgres database, REDIS cache, and the qlever (Bast & Buchhold, 2017) graph database.

4.4 Gut-Brain Viewer

We also deployed the gut-brain viewer by SURF on the *Demo System* under <https://hereditary.cgv.tugraz.at/gut-brain-viewer>. A screenshot of the tool is shown in fig. 5.

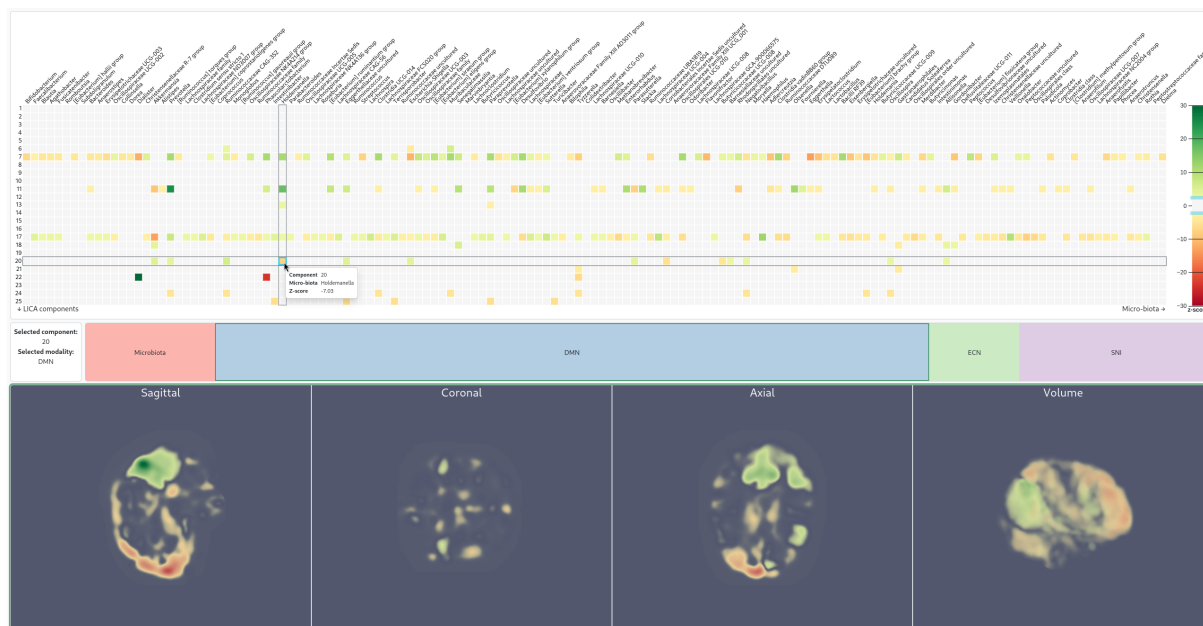


Figure 5: Screenshot of the currently deployed Gut-Brain Viewer.

4.5 AlloyInter

One of the most recently developed components, working title AlloyInter (Kantz, Waldert, et al., 2025), visualizes high-dimensional parameter space data. The approach is useful, e.g., for simulation data, when (1) multiple parameters control the simulation, and (2) the output of the simulation is described again by multiple measures. Specifically, we apply dimensionality reduction to visualize the simulation output, and use glyphs to show the input. Interaction allows us to relate the input and output spaces and choose from several outcomes. To test and start the approach, we obtained a scientific dataset from materials science, focusing on alloy mixtures and their properties. While these are not from the biomedical domain, the component has been evaluated and is ready for application to use cases in HEREDITARY. The tool is, similar to the others, archived using Zenodo Kantz, 2025a and deployed on the *Demo System* under <https://hereditary.cgv.tugraz.at/alloy-inter>. A screenshot of the system can be found in fig. 6.

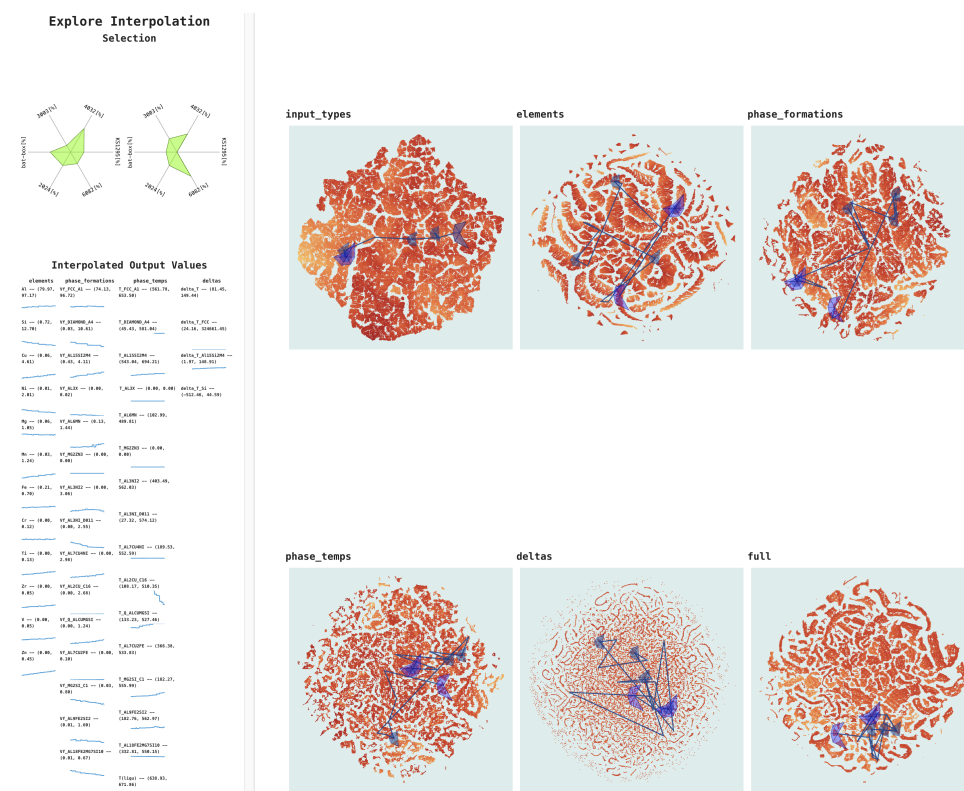


Figure 6: Screenshot of the currently deployed AlloyInter tool.

5 Demonstration Video Material

To make applications known and inform collaborators about their usage possibilities, we produced demonstration videos for several important software components and applications.⁴ fig. 7 gives an overview of the videos. Specifically:

- **Clusters in Focus:** Comparison of clusters for feature selection, e.g., for biomarker identification, and phenotypization applications (cf. Section 4.1).
- **Neurodegen-Vis:** Correlation analysis of features in biomedical data analysis, supported by an LLM-based chat assistant for exploration and guidance (Neurodegen-Vis; cf. Section 4.2).
- **OnSET:** Visual exploration and searching in knowledge graph data (OnSET; cf. Section 4.3).
- **Gut-Brain Viewer:** Linked visual exploration of Magnetic Resonance Imaging (MRI) imaging data and gut microbiota data, from linked component analysis (cf. Section 4.4).
- **Droplets:** Glyph design for overlaying multidimensional descriptive data onto regions of medical images. This visualization technique is intended to be integrated into other tools, not a standalone tool; and therefore only shown as a demonstration video (Schreck, Lengauer, Waldert, Kantz, Grabner, Schilcher, Percic, & van Leeuwen, 2024).

For details on these components, please see Deliverables D5.1, D5.3, and D5.5 (Schreck, Lengauer, Waldert, Kantz, Grabner, Schilcher, Percic, & van Leeuwen, 2024; Schreck, Lengauer, Waldert, Kantz, Grabner, Schilcher, Percic, van Leeuwen, et al., 2024; Schreck et al., 2025). It is planned to release the videos through the HEREDITARY YouTube channel and promote them in our network. Pre-release versions of the videos are available on <https://cloud.tugraz.at/index.php/s/C939k8sXo9WwtnN>.

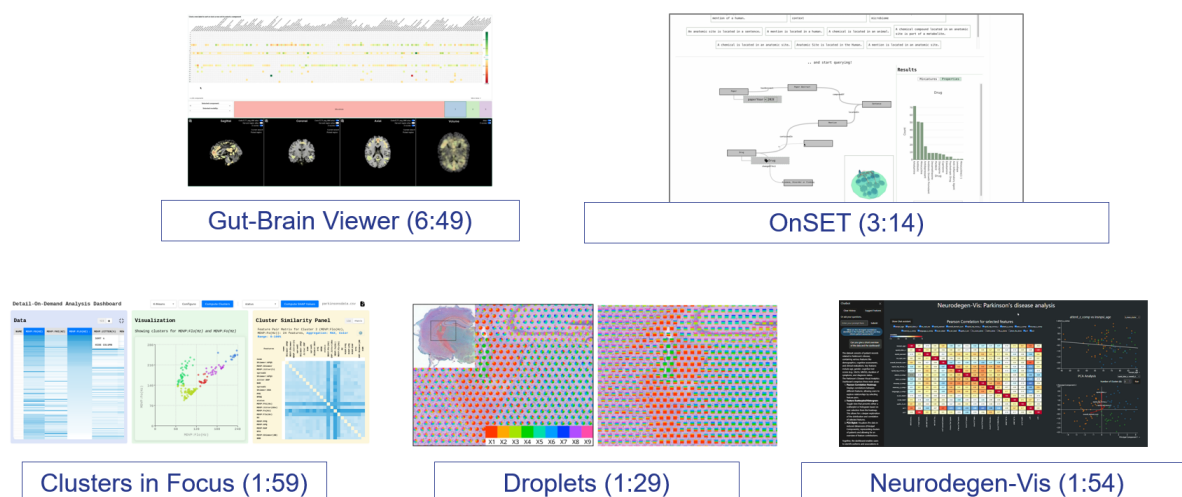


Figure 7: Demonstration videos for main components developed in the project.

⁴AlloyInter is not part of this collection due to the breaking nature of the work, and the tool still being in development.

6 Conclusion and Next Steps

In this work, we introduced demonstrators for the visualisation components in the Hereditary project. Building on visualization components, we gradually develop applications that are deployed on a dedicated demonstration server. We provide our components and applications following an open science approach, corresponding to the deliverable being a PU (public) release. Demonstration videos exemplify the components. The components and application development is informed by activities in requirement engineering and evaluation (see D5.7). Future work includes extending the components, building applications, and applying them to HEREDITARY analysis cases in different domains. We have already implemented some form of user guidance in our approaches, e.g., the LLM-based explanation function in the Neurodegen-Vis application (deployment forthcoming). The further developments are informed by activities in requirement analysis and researching user guidance approaches.

References

- Bast, H., & Buchhold, B. (2017). Qlever: A query engine for efficient sparql+text search. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 647–656. <https://doi.org/10.1145/3132847.3132921>
- FAIRsharing Team. (2024). Fairsharing record for: Rdf/xml syntax specification. <https://doi.org/10.25504/FAIRSHARING.5CAE72>
- FAIRsharing Team & McQuilton, P. (2024). Fairsharing record for: Comma-separated values. <https://doi.org/10.25504/FAIRSHARING.1943D4>
- Grabner, M., Waldert, P., & Kantz, B. (2025, December). *Hereditary-eu/neurodegen-vis: Initial release* (Version v0.1.0). Zenodo. <https://doi.org/10.5281/zenodo.17870087>
- Grabner, M., Waldert, P., Kantz, B., & Schreck, T. (2025, October). Privacy-Preserving Visual Exploration of Healthcare Data with a Language Model Integration. In *Joint Proceedings of the ECAI 2025 Workshops XAI-KRKG and UCEX-XAI, Bologna, Italy*. <https://tugraz.elsevierpure.com/en/publications/privacy-preserving-visual-exploration-of-healthcare-data-with-a-l>
- Kantz, B. (2025a, October). *Dakantz/alloyinter: Vis 2025* (Version vis-2025). Zenodo. <https://doi.org/10.5281/zenodo.17415354>
- Kantz, B. (2025b, November). *Hereditary-eu/onset: V0.3.0-alpha* (Version v0.3.0). Zenodo. <https://doi.org/10.5281/zenodo.17606868>
- Kantz, B., Innerebner, K., Waldert, P., Lengauer, S., Lex, E., & Schreck, T. (2025a). Graph queries from natural language using constrained language models and visual editing. *Proceedings of the 2024 IEEE International Conference on Knowledge Graph (ICKG)*.
- Kantz, B., Innerebner, K., Waldert, P., Lengauer, S., Lex, E., & Schreck, T. (2025b). Onset: Ontology and semantic exploration toolkit. *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 3980–3984. <https://doi.org/10.1145/3726302.3730148>
- Kantz, B., Waldert, P., Lengauer, S., & Schreck, T. (2025). Alloyinter: Visualising alloy mixture interpolations in t-sne representations. <https://arxiv.org/abs/2509.19202>
- Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P. A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J. L., Chue Hong, N., Goble, C., & Capella-Gutierrez, S. (2019). Towards FAIR principles for research software. *Data Sci.*, 3(1), 37–59. <https://doi.org/10.3233/DS-190026>
- Laurent, A. M. S. (2004). *Understanding open source and free software licensing: Guide to navigating licensing issues in existing & new software*. " O'Reilly Media, Inc."
- Peters, I., Kraker, P., Lex, E., Gumpenberger, C., & Gorraiz, J. I. (2017). Zenodo in the Spotlight of Traditional and New Metrics. *Front. Res. Metrics Anal.*, 2, 318745. <https://doi.org/10.3389/frma.2017.00013>
- Schilcher, L., & Waldert, P. (2025, November). *Hereditary-eu/clustersinfocus: Initial release* (Version v0.1.0). Zenodo. <https://doi.org/10.5281/zenodo.17661296>
- Schilcher, L., Waldert, P., Kantz, B., & Schreck, T. (2025a). Clusters in Focus: A Simple and Robust Detail-On-Demand Dashboard for Patient Data. In L. Garrison & R. Krueger (Eds.),

- Eurographics workshop on visual computing for biology and medicine*. The Eurographics Association. <https://doi.org/10.2312/vcbm.20251250>
- Schilcher, L., Waldert, P., Kantz, B., & Schreck, T. (2025b, October). *Clusters in Focus: A Simple and Robust Detail-On-Demand Dashboard for Patient Data* (Version 1.0.0). <https://github.com/hereditary-eu/ClustersInFocus>
- Schreck, T., Lengauer, S., Waldert, P., Kantz, B., Boytcheva, S., Datseris, A., Koleva, T., Primov, T., Nikolova-Koleva, I., Rodríguez, J. M., & Silvello, G. (2025, August). Deliverable 5.5: Querying components. <https://doi.org/10.5281/zenodo.17368336>
- Schreck, T., Lengauer, S., Waldert, P., Kantz, B., Grabner, M., Schilcher, L., Percic, D., & van Leeuwen, C. (2024, December). Deliverable 5.3: Visualization components for spatial, image, and simulation data. <https://doi.org/10.5281/zenodo.14628150>
- Schreck, T., Lengauer, S., Waldert, P., Kantz, B., Grabner, M., Schilcher, L., Percic, D., van Leeuwen, C., Lissandrini, M., & Romanovych, A. (2024, December). Deliverable 5.1: Visualization components for sequences, networks, text, and high- dimensional data. <https://doi.org/10.5281/zenodo.14628086>