

Genetic networks: data model and visualisations

Spadini, Elena

elena.spadini@unibas.ch
Universität Basel, Switzerland

Christen, Alessio

alessio.christen@unil.ch
Université de Lausanne, Switzerland

Pallacci, Valentina

valentina.pallacci@mail.polimi.it
Politecnico di Milano, Italy

Elli, Tommaso

tommaso.elli@polimi.it
Politecnico di Milano, Italy

Benedetti, Andrea

andrea.benedetti@polimi.it
Politecnico di Milano, Italy

Maggetti, Daniel

Daniel.Maggetti@unil.ch
Université de Lausanne, Switzerland

Mauri, Michele

michele.mauri@polimi.it
Politecnico di Milano, Italy

Pétermann, Stéphane

stephane.petermann@unil.ch
Université de Lausanne, Switzerland

In this paper, we introduce the concept of genetic networks to study the relationships between manuscripts and publications in genetic criticism and scholarly editing. We present a data model and its formalisation in an OWL2 ontology, as well as the corresponding data visualisations in the form of sky maps. Our case study is the literary œuvre of the Swiss writer and photographer Gustave Roud (1897-1976). We conclude with some remarks about the project workflow and the integration of data and visual modelling.

Genetic Networks

In the fields of genetics criticism and digital scholarly editing, modelling efforts have focused primarily on the information contained in a single document and on the comparison between documents. The Text Encoding Initiative (TEI) provides a suitable architecture to track and digitally represent the writing process happening on the page.

The concept of genetic network complements the modelling of single documents and addresses instead the relationships between them. The object of study shifts to the clustering of documents in genetic dossiers and their connection to publications. This is not something new nor specific to the digital medium: ordering materials is a crucial step when studying modern manuscripts, corresponding to the *classement des manuscrits* phase in the six steps identified by De Biasi (2008). Daniel Ferrer too gave voice to the need of studying the relationships between documents: reporting on experiments on how to place and link the genetic documents in the hyperspace, he wrote: "Ce qui est sans doute à retenir de cette expérience [with the software Storyspace] [...] c'est la nécessité de nommer les liens et de prévoir des carrefours de liens" (Ferrer 2008, italics in the original).

The novelty of our proposal is to make explicit and computable the ordering of the documents through data modelling and visualisation, instead of it being a necessary but implicit step in the study of genetic materials. Furthermore, we propose the concept of network to complement and broaden that of dossier: multiple genetic dossiers may be intertwined, for example when the same diary note is reused in several writing projects or when a publication is reused and becomes part of the genetic materials for a later work.

Gustave Roud

The genesis of the literary works of the Swiss author Gustave Roud can be well represented as genetic networks. His writing practices are based on the selection and reuse of existing and sometimes previously published materials, a process that cannot be traced by looking at a single manuscript, nor at the variants between two or more witnesses.

This was one of the challenges posed by the project "Gustave Roud. Œuvres complètes" (University of Lausanne, 2017-2022, funded by the Swiss National Science Foundation). The aim of the project has been to produce a print edition of his complete works (Gustave Roud, *Œuvres complètes*, dir. Claire Jaquier et Daniel Maggetti, Genève, Zoé, 2022), complemented by a web application to access the drafts and manuscripts (<https://roud.unil.ch>).

The project data is stored at the Swiss National Data and Service Centre for the Humanities (DaSCH), which provides DSP (DaSCH Service Platform), a RDF-based repository for the long-term preservation and reuse of data.

The GENO ontology

As part of the project, the OWL2 GENO ontology for representing genetic networks was developed (see <https://gen-o.github.io/>). In the case of Gustave Roud, modelling the data with GENO allowed us to organise a considerable amount of genetic materials enriched with a semantic structure: 202 avant-textual documents (including lists, drafts, clean versions, etc.) and 259 diary entries, linked to 93 genetic dossiers. It therefore becomes possible to make queries concerning one or more genetic dossiers and Roud's writing processes, including his reuse practices: for example, which works are rooted in the diary? Which avant-textual materials belong to a specific dossier? In what context is a specific article reused?

The first version of GENO was described by Christen and Spadini (2019) and the second version will be presented at the conference. The new version takes into account endogenetic and exogenetic materials, uses inferences to distinguish composition phases,

allows to differentiate the genesis of different sections of a work, and integrates some FRBR concepts. It should be suitable to represent a larger variety of literary genesis and accommodate different levels of description.

Data Model and Data Visualisations

Once the data was modelled according to the GENO ontology, it soon became clear that we ought to find suitable visualisations to represent these complex genetic networks. To this end, the researchers involved in the project collaborated with designers from DensityDesign Lab, a laboratory of the Politecnico in Milan (<https://densitydesign.org/>), who created data visualisations of the genetic networks of Roud's most important publications in the form of sky maps (available at <https://roud.unil.ch>; one of them is in Fig. 1).

The chosen method was participatory design, which resulted in a stimulating and satisfying collaboration for both parts. The design process included definition of design requirements, prototyping and evaluation, implementation of artefacts, and final assessments (Elli et al. in press).

The JSON-LD data for the visualisations were extracted from the full project dataset using SPARQL queries (see <https://github.com/gustaveroudproject/geneticNetworksDataViz>) and part of the implementation was conducted using Observable (<https://observablehq.com/@densitydesign/snodaroud>), which facilitated communication and allowed inconsistencies in the data to be detected.

The visualisations confirm recurring structures among networks, corresponding to Roud's creative habits, such as the "marionette" structure (see, for example, *Essai pour un paradis*). Furthermore, they make clear how sometimes diary notes from the same notebook are reused in different articles and then reunited when the articles are repurposed in the genesis of a single book (*Feuillets and Campagne perdue*).

The experience of the project taught us that the development of a visual model can be part of the development of a data model: the collaboration with designers and the design process then add to the iterative nature of the data modelling process itself.

Conclusions

In this paper we propose the concept of literary genetic networks and we present an ontology to model them in the context of genetic editing (GENO), as well as a corresponding visualisation model. Both the ontology and the visual model can be reused, to make explicit and computable the ordering of genetic materials in other literary genesis. The workflow too, integrating data and visual modelling, might be relevant for other projects.

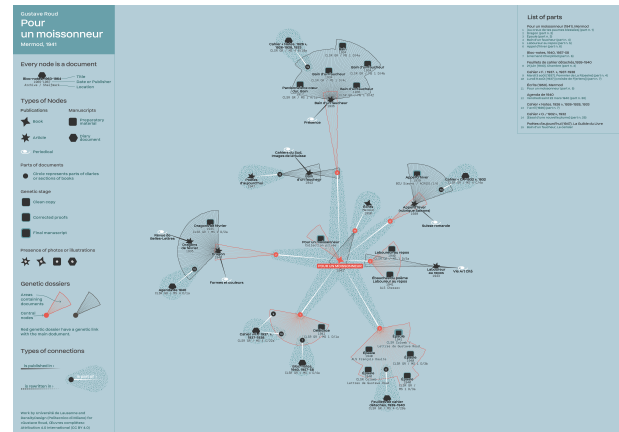


Figure 1: Genetic network of Gustave Roud, *Pour un moissonneur*, 1941.

Bibliography

Christen, Alessio / Spadini, Elena (2019): "Modeling genetic networks. Gustave Roud's œuvre, from diary to poetry collections", in: *Umanistica Digitale* 7. DOI: 10.6092/issn.2532-8816/9063.

De Biasi, Pierre-Marc (2008): "Les six grandes étapes de la recherche en génétique des textes", in: Crasson, Aurèle (ed.): *L'édition du manuscrit. De l'archive de création au scriptorium numérique*. Louvain-La-Neuve: Bruylant-Academia, 25–46.

Ferrer, Daniel (2008): "La représentation hypertextuelle des manuscrits : quelques leçons de douze années d'expériences", in: Crasson, Aurèle (ed.): *L'édition du manuscrit. De l'archive de création au scriptorium numérique*. Louvain-La-Neuve: Bruylant-Academia, 189–208.

Elli, Tommaso / Benedetti, Andrea / Pallacci, Valentina / Spadini, Elena / Mauri, Michele (in press): "Designing network visualizations for genetic literary criticism", in *Convergências* 16.