Structuring and Analyzing Oral and Dental Data from the GenIDA Database in Rare Diseases with Intellectual Disability

Djawad Hachemaoui[1,2], Victoria Hinckelmann[3], Jean-Louis Mandel[3], Agnès Bloch-Zupan[1,2,3,4] Alexandra Jimenez-Armijo[2,3]

1.University of Strasbourg, Faculty of Dental Surgery, Strasbourg, France
2.Strasbourg University Hospitals (HUS), Department of Medicine and Oral Surgery, Civil Hospital, Reference Center for Rare Oral and Dental Diseases (O-Rares), Rare Diseases Health Network TETECOU, European Reference Network ERN CRANIO, Strasbourg, France
3.University of Strasbourg, Institute of Genetics and Molecular and Cellular Biology (IGBMC), INSERM U1258, CNRS UMR7104, Illkirch, France
4.University of Strasbourg, Institute for Advanced Study (USIAS), Strasbourg, France

INTRODUCTION

Rare genetic diseases frequently display oro-dental manifestations that may represent important diagnostic or prognostic indicators.
However, these data are often poorly structured and difficult to exploit within existing clinical databases.

The GenIDA database provides a unique opportunity for international collection of phenotypic data, but it requires prior methodological structuring to enable meaningful scientific analysis.

The GenIDA Database

GenIDA (Genetic Intellectual Disability Database) is an international participatory database dedicated to genetic forms of intellectual disability, with or without autism spectrum disorders or epilepsy.

Its objective is to better characterize clinical manifestations and the natural history of these rare diseases (Figure 1), which remain insufficiently described due to their low prevalence.

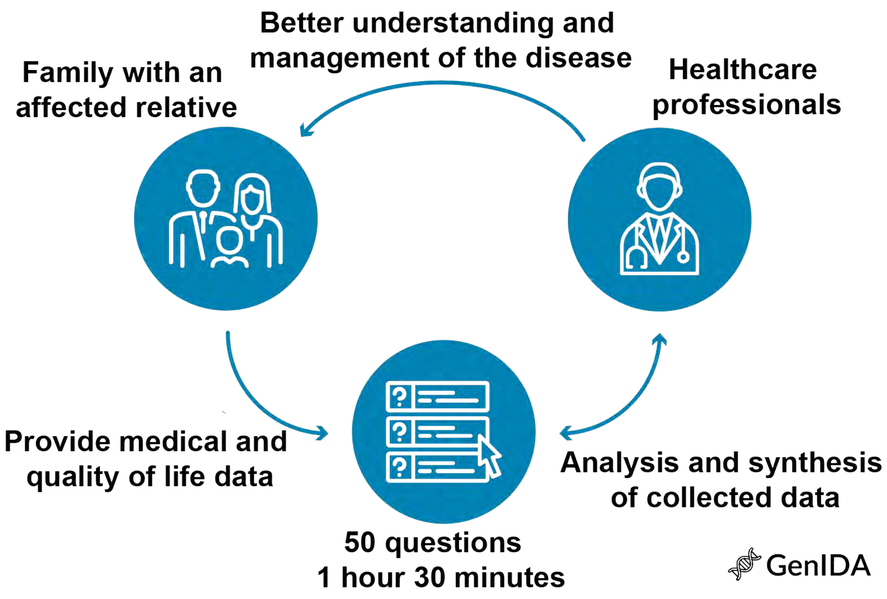

Figure 1. Overview of the GenIDA database workflow

Problem Statement

Oro-dental information in GenIDA:

- is scattered throughout the questionnaire and organized into dependent questions
- includes numerous free-text responses written by parents in non-medical language (Figure 2)
- originates from multiple countries and languages, requiring prior harmonization (Figure 2)

This organization makes direct data exploitation complex and limits their use for clinical research and scientific publications.

Objectives

- To make oro-dental data from the GenIDA database usable for clinical research
- To develop a reproducible methodology applicable to all rare diseases described in the GenIDA database

METHODOLOGY

The work is based on a progressive approach combining data exploration and methodological structuring.

Database Familiarization

- Understanding the structure of the GenIDA database (table organization, relationships between variables, observation levels)
- Identification of oro-dental variables
- Selection of rare diseases of interest

Data Structuring

- Extraction and identification of oro-dental-related questions
- Data cleaning: handling missing values and correcting inconsistencies
- Automatic association of responses linked to the same clinical entity
- Data processing and structuring using Python programming language (pandas)

Translation and Clinical Standardization

- Translation of responses from international questionnaires
- Processing of free-text responses written by parents
- Standardization of clinical oro-dental descriptions using the HPO ontology (Human Phenotype Ontology)

Descriptive and Exploratory Analyses

- Analysis of frequencies and oro-dental clinical profiles
- Cross-disease comparisons
- Interpretation of results in relation to the scientific literature
- Descriptive analyses and visualizations using R / RStudio software (tables and graphs)

Data Exploitation and Output

- Final structuring of data into a format usable for clinical research
- Proposal of a reproducible methodology applicable to other rare diseases
- Use of Excel tables for visualization and presentation of results

APPLICATION & PERSPECTIVES

Gender	Age ↑	Responses
Man	8	weinig glazuur
Woman	9.8	Not all baby teeth have come in yet
Woman	13	super pointy, fine canine teeth
Man	2.3	retard d'évolution dentaire
Man	17	Still has some baby teeth, yet getting wisdom teeth too
Man	5.1	Still has some baby teeth, yet getting wisdom teeth too
Man	5.1	gaps between the teeth

Figure 2. Examples of free-text responses describing oro-dental anomalies reported by families in the GenIDA database, illustrating the linguistic and clinical heterogeneity of the data.

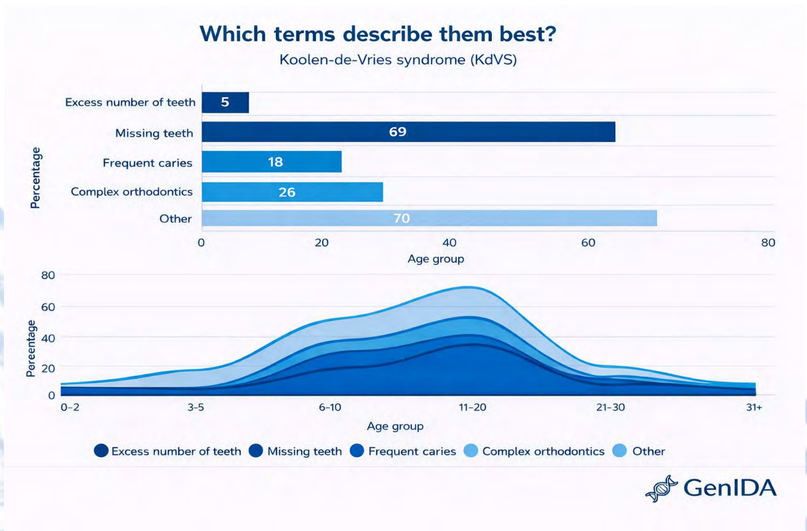

Figure 3. Distribution of oro-dental anomalies reported in patients with Koolen-de Vries syndrome according to age (GenIDA data).

Application Case: Koolen-de Vries Syndrome

Koolen-de Vries syndrome is a rare genetic disorder well represented in the GenIDA database (Figure 3).

It provides a relevant example to:

- test the data structuring methodology
- identify the available oro-dental data
- illustrate the clinical relevance of the proposed approach

Perspectives

The structuring of oro-dental data from the GenIDA database will enable:

- better clinical use of the data
- identification of clinical signs and oro-dental features that are still poorly described or absent from the literature
- extension of the methodology to other phenotypic domains
- development of descriptive and comparative analyses
- facilitation of research work and scientific publications

CONCLUSION

This project aims to render complex oro-dental data from an international registry usable for research.

The proposed methodology provides a reproducible framework that can support clinical research and the scientific use of GenIDA data.

djawad.hachemaoui@etu.unistra.fr