Rajzvizsgálati adatbázis / Projective Drawing Database

Vass, Z. (2025). *Rajzvizsgálati adatbázis / Projective Drawing Database* (Version 1.0) [Database]. Károli Gáspár Reformed University. https://krepozit.kre.hu/

Database Documentation
Last updated: 2025-10-10

What this database is

Rajzvizsgálati adatbázis (Projective Drawing Database) – a comprehensive methodology for building thematic databases of projective drawing analysis from validated case studies.

- **Timeframe**: January 6, 2025 - May 15, 2025
- **Delivery format**: Multi-component digital system including:
  - Searchable JSON database (`patterns.json`)
  - Interactive HTML search interface (`search_interface.html`)
  - RAG (Retrieval-Augmented Generation) search engine
  - LLM-based pattern extraction pipeline
  - Comprehensive sign indexing system
- **Source materials**: Case studies and professional documentation collected over five years
- **Author and affiliation**: see Metadata below

What the database contains (methodology and schema)

Core Components

The database methodology encompasses the following structured units:

1. **Validated, anonymized case studies**
   - Full compliance with GDPR requirements
   - Complete anonymization protocols
   - Professional-ethical guidelines for data usage
2. **Diagnostically instructive case collections**
   - Educationally valuable case presentations
   - Rare or exceptional drawing assessment phenomena
   - Typical case demonstrations for teaching purposes
   - Sample analyses supporting diagnostic practice
3. **Methodological framework**
   - Structured template for drawing assessment documentation
   - Systematic pattern extraction procedures
   - Sign co-occurrence analysis protocols
   - Confidentiality requirement adherence guidelines

Database Schema (patterns.json)

Top-level JSON structure:

- **metadata**: object
  - title: string ("Rajzvizsgálati adatbázis / Projective Drawing Database")
  - author: string ("Vass, Zoltan")
  - year: string ("2025")
  - version: string ("1.0")
  - publisher: string ("Károli Gáspár Reformed University")
  - location: string ("Budapest, Hungary")
  - url: string ("https://krepozit.kre.hu/")
  - citation: string (APA 7 formatted)
  - total_source documents: number (currently partially processed)
  - total_patterns: number (currently 240)
  - total_reference_signs: number (1,116 indexed signs)
  - categories: array (11 sign categories)
- **patterns**: object<string, object>
  - Key: pattern ID (normalized combination of sign IDs, e.g., "20_146_996")
  - Value: object containing:
    - sign_ids: array (IDs of co-occurring signs)
    - sign_names: array (descriptive names of signs)
    - occurrences: array (source document IDs where this pattern appears)
    - interpretations: array (contextual interpretations from source documents)
    - frequency: number (count of occurrences)
- **comprehensive_sign_index**: object<string, object>
  - Key: sign ID as string (e.g., "990", "1087", "1112")
  - Value: object containing:
    - id: number (unique sign identifier from reference file)
    - name: string (sign description in Hungarian)
    - category: string (one of 11 categories: ANAMNÉZIS, EMBERRAJZ, ÁLLATRAJZ, FARAJZ, HÁZRAJZ, SZABADRAJZ, AKTUÁLGENEZIS, INTUITÍV ELEMZÉS, EGÉSZLEGES ELEMZÉS, FORMAI-SZERKEZETI JELLEMZŐK, TARTALMI-SZIMBOLIKUS JELLEMZŐK)
    - patterns: array (pattern IDs containing this sign)
    - source documents: array (source document IDs mentioning this sign)

Reference Sign Categories (11 categories)

1. **ANAMNÉZIS** (Anamnestic Data)
   - Demographic information
   - Psychological diagnosis
   - Social behavior
   - Emotional characteristics
   - Cognitive functioning

- Somatic diseases

2. **AKTUÁLGENEZIS** (Process Analysis)
    - Verbal behavior during testing
    - Gestures and facial expressions
    - Test-taking behavior
    - Psychomotor characteristics
    - Drawing time and reaction time
    - Drawing sequence
    - Erasing patterns
3. **INTUITÍV ELEMZÉS** (Intuitive Analysis)
    - Receptive observation methods
    - Spontaneous attention focus
    - Motor empathy
    - Kinesthetic empathy
    - Empathic questioning
4. **EGÉSZLEGES ELEMZÉS** (Holistic Analysis)
    - Emotional-mood tone
    - Integration and harmony
    - Hárdi drawing personality levels
    - Movement and form imagery
    - Color usage
    - Elkisch analysis framework
5. **FORMAI-SZERKEZETI JELLEMZŐK** (Formal-Structural Characteristics)
    - Size and proportions
    - Symmetry
    - Position on page
    - Line quality
    - Shading
    - Detail level
    - Transparency
6. **TARTALMI-SZIMBOLIKUS JELLEMZŐK** (Content-Symbolic Characteristics)
    - Eleven areas of content-symbolic analysis
    - Iconic identification of motifs
    - Semantic connections and interactions
    - Cultural and social influences
    - Personal relationships
    - Psychoanalytic symbol analysis
7. **HÁZRAJZ** (House Drawing)
    - Thematic stereotypes

- Door characteristics
- Windows
- Roof and chimney
- Walls
- Stairs and pathways
- Environment and accessories

8. **FARAJZ** (Tree Drawing)
    - Holistic tree characteristics
    - Thematic stereotypes
    - Tree species
    - Foliage and branches
    - Tree trunk
    - Roots
    - Hollow
    - Special markings
9. **ÁLLATRAJZ** (Animal Drawing)
    - Animal species categories
    - Movement and posture
    - Head and features
    - Body trunk
    - Limbs
    - Latent animal imagery
10. **EMBERRAJZ** (Human Figure Drawing)
    - Gender representation
    - Age characteristics
    - Thematic stereotypes
    - Posture and movement
    - Head and neck
    - Facial features
    - Body trunk
    - Arms, hands, fingers
    - Legs, feet, toes
    - Hair and body hair
    - Shoulders
    - Chest
    - Hip and genital area
    - Clothing and accessories
11. **SZABADRAJZ** (Free Drawing)
    - Thematic stereotypes
    - Typical representation forms

- Accessory elements (ground line, background, objects, natural elements)

Current statistics (from patterns.json)

- **Total reference signs**: 1,116 (all indexed and searchable)
- **Processed source documents**: (654 available, currently partially processed)
- **Extracted patterns**: 240 (co-occurrence patterns of 3+ signs)
- **Signs appearing in patterns**: 20 signs
- **Signs searchable but not yet in patterns**: 1,096 signs
- **Sign categories**: 11 major categories
- **Database file size**: approximately 540 KB (JSON format)

Methodology and processing pipeline

1. Source Data Collection

Input: CSV file containing source document data with the following fields:

- Title (Hungarian)
- Summary/Abstract (Hungarian)
- Analysis/Discussion (Hungarian)
- Metadata (year, author, etc.)

Current dataset: 654 source documents available for processing

2. LLM-Based Sign Extraction

Technology stack:

- Extraction method: Structured prompt-based extraction
- Reference library: 1,305 unique sign IDs from reference file
- Language: Hungarian (all processing in native language)

Extraction process (`DrawingPatternExtractor`):

1. Load source document text (summary + analysis)
2. Submit to LLM with reference sign library
3. Extract:
   - Identified sign IDs mentioned in text
   - Contextual interpretations
   - Demographics (age, gender, occupation, anamnestic features)
4. Validate extracted sign IDs against reference library
5. Cache results to avoid re-processing

Batch processing (`PatternDatabaseBuilder`):

- Processes source documents in configurable batches (default: 10)
- Manual checkpoints between batches
- Progress tracking with visual indicators

- Extraction cache for efficiency

3. Pattern Discovery

Co-occurrence analysis:

- Minimum pattern size: 3 signs appearing together
- Pattern identification: Normalized combination of sign IDs
- Frequency tracking: Count of occurrences across source documents
- Interpretation aggregation: Collect contextual meanings

Example pattern:

```
{
  "20_146_996": {
    "sign_ids": [20, 146, 996],
    "sign_names": [
      "alacsony szocioökonómiai státus",
      "kényszeres személyiségvonások",
      "fejábrázolás egyszerű körsémával"
    ],
    "occurrences": ["source document_1", "source document_3"],
    "interpretations": [
      "Simple drawing style correlating with socioeconomic factors",
      "Obsessive-compulsive traits reflected in controlled circular forms"
    ],
    "frequency": 2
  }
}
```

4. Comprehensive Sign Indexing

Reference library parsing (`build_comprehensive_sign_index.py`):

1. Parse `drawing-and-anamnestic-signs-features.txt`
2. Extract 1,116 unique signs with categories
3. Assign sequential IDs (1-1305, with gaps for hierarchical structure)
4. Index all signs regardless of source document occurrence
5. Link signs to patterns (when applicable)
6. Link signs to source documents (when mentioned)

Result: Every reference sign is searchable, whether or not it has been found in processed source documents yet.

5. Search Engine

Dual search functionality (`RAGSearchEngine`):

Pattern search (`search_patterns()`):

- Searches across 240 co-occurrence patterns
- Returns matches with full source document context
- Includes interpretations and frequency data

Sign search (`search_signs()`):

- Searches all 1,116 reference signs
- Keyword-based matching (sign name, category)
- Relevance scoring (exact match > partial match > keyword match)
- Returns sign metadata + pattern/source document links

Example searches:

- `ujj` (finger) → finds 10+ finger-related signs
- `határozott` (determined) → finds 3 patterns + 5 individual signs
- `FARAJZ` (tree drawing) → finds all tree-related signs
- `anamnézis` → finds anamnestic data signs

6. Interactive HTML Interface

Features (`search_interface.html`):

- Real-time search with dual result display
- Pattern matches section (with source document examples)
- Individual sign matches section (all reference signs)
- Example query buttons for quick exploration
- Statistics dashboard
- Responsive design for readability

Technology:

- Single-file HTML with embedded JavaScript
- Fetches data from JSON via local server
- Client-side search and filtering
- No external dependencies

7. Deployment

Local server (`serve_search.py`):

```
python3 serve_search.py
# Opens browser at: http://localhost:8000/search_interface.html
```

Server features:

- Simple HTTP server for local development
- Auto-opens browser on start
- Serves JSON database and HTML interface
- CORS-enabled for local testing

Design principles and quality assurance

Professional-Ethical Framework

1. **Anonymization protocols**:
    - Complete removal of personal identifiers

- GDPR-compliant data handling
- Secure storage and access controls

2. **Confidentiality requirements**:
   - Professional use only (research, education)
   - No distribution of raw case data
   - Aggregated patterns for public presentation
3. **Data validation**:
   - Expert review of extracted patterns
   - Cross-validation with reference literature
   - Manual verification of diagnostic correlations

Technical Quality Standards

1. **LLM extraction accuracy**:
   - Reference library validation (all sign IDs checked)
   - Contextual interpretation preservation
   - Hungarian language processing fidelity
2. **Database integrity**:
   - Unique pattern identification
   - Consistent source document ID mapping
   - Traceability to source data (CSV row numbers)
3. **Search performance**:
   - Keyword-based relevance scoring
   - Fast client-side filtering
   - Comprehensive coverage (all 1,116 signs)
4. **Extensibility**:
   - Modular pipeline design
   - Incremental processing capability
   - Merge-friendly database structure

Methodological documentation template

The database provides a **structured template for drawing assessment documentation**, based on the eleven-category framework from `drawing-and-anamnestic-signs-features.txt`:

I. Anamnestic Data (ANAMNÉZIS)

- Demographics: age, gender, occupation
- Psychological diagnosis (DSM-IV, BNO-10)
- Personality characteristics
- Social behavior patterns

- Cognitive functioning
- Somatic conditions

II. Process Analysis (AKTUÁLGENEZIS)

- Verbal behavior during testing
- Gestures and facial expressions
- Test-taking behavior
- Psychomotor characteristics
- Temporal features (drawing time, reaction time)
- Drawing sequence and modifications
- Erasing patterns

III. Intuitive Analysis (INTUITÍV ELEMZÉS)

- Initial impressions
- Spontaneous attention focus
- Motor and kinesthetic empathy
- Visualization techniques
- Empathic questioning

IV. Holistic Analysis (EGÉSZLEGES ELEMZÉS)

- Emotional-mood tone
- Integration and harmony levels
- Drawing personality levels (Hárdi framework)
- Movement and form emphasis
- Color usage patterns
- Elkisch analysis dimensions

V. Formal-Structural Characteristics (FORMAI-SZERKEZETI JELLEMZŐK)

- Size and proportions
- Symmetry features
- Position on page
- Line quality
- Shading techniques
- Detail level
- Transparency effects

VI. Content-Symbolic Characteristics (TARTALMI-SZIMBOLIKUS JELLEMZŐK)

- Iconic identification
- Semantic connections
- Cultural influences
- Personal relationships
- Symbol analysis (archetypal, cultural, individual)

- Psychoanalytic interpretations

VII-XI. Specific Drawing Types

- **HÁZRAJZ** (House Drawing): architectural features, symbolic meanings
- **FARAJZ** (Tree Drawing): vitality, growth, environmental connection
- **ÁLLATRAJZ** (Animal Drawing): instinctual aspects, projected traits
- **EMBERRAJZ** (Human Figure Drawing): self-concept, body image, identity
- **SZABADRAJZ** (Free Drawing): spontaneous expression, personal themes

Future directions and scalability

Expansion potential

1. **Full dataset processing**:
    - Current: 13 source documents processed
    - Target: 654 source documents
    - Estimated time: 6-10 hours of LLM processing
    - Estimated cost: $15-30 in Gemini API fees
    - Expected result: 200-500 signs appearing in patterns
2. **Enhanced pattern types**:
    - 2-sign co-occurrences (currently minimum 3)
    - Sequential patterns (drawing process order)
    - Conditional patterns (if X then Y)
    - Diagnostic category correlations
3. **Advanced analytics**:
    - Pattern frequency distribution analysis
    - Sign clustering by co-occurrence
    - Diagnostic predictive modeling
    - Visualization of sign networks

Methodological refinements

1. **LLM extraction improvements**:
    - Fine-tuned prompts for better sign recognition
    - Multi-pass extraction for recall improvement
    - Human-in-the-loop validation workflows
2. **Search enhancements**:
    - Semantic similarity search (beyond keywords)
    - Pattern recommendation engine
    - Diagnostic query assistance
3. **Educational integration**:

- Case study presentation tools
- Interactive learning modules
- Assessment practice scenarios

Usage guidelines

For researchers

Accessing the database:

1. Clone repository from KREPOZIT
2. Install dependencies: `pip install -r requirements.txt`
3. Start local server: `python3 serve_search.py`
4. Open browser at `http://localhost:8000/search_interface.html`

Search strategies:

- **Exploratory**: Use broad keywords (e.g., "szorongás" for anxiety-related signs)
- **Specific**: Use exact sign names (e.g., "lábujjhegyen állás")
- **Category-based**: Use category names (e.g., "FARAJZ" for all tree signs)
- **Pattern-based**: Use interpretive terms (e.g., "határozott" for determined patterns)

Interpreting results:

- **Pattern matches**: Show co-occurring signs with source document examples
- **Sign matches**: Show individual signs with metadata
- Signs marked "Még nem jelenik meg feldolgozott mintázatokban" are searchable but not yet found in processed source documents

For educators

Teaching applications:

1. **Case study exploration**: Search for specific diagnostic patterns
2. **Sign familiarization**: Browse all 1,116 reference signs by category
3. **Pattern recognition training**: Compare patterns across source documents
4. **Documentation practice**: Use template for student assessments

Ethical considerations:

- All case data is anonymized
- Use only for educational purposes
- Maintain confidentiality in all contexts

For processing new data

Adding more source documents:

Option 1: Process specific number (e.g., 10 source documents):

```
# Create process script
```

```
python3 process_10_more_source documents.py
```

Option 2: Process all remaining (654 total):

```
USE_EXTERNAL_LLM=1 python3 build_rag_database.py
```

Incremental processing:

- New source documents are merged with existing database
- source document IDs are auto-incremented
- Patterns are deduplicated and merged
- Sign index is updated with new links

Quality checks:

- Verify sign extraction (check for "⚠️ Nem talált jeleket" warnings)
- Review pattern frequency (minimum 3 signs per pattern)
- Validate source document text length (should be > 100 chars)

Technical architecture

File structure

```
kre-palyazat-rajzvizsgalati-adatbazis/
├── rag_system/
|   ├── data/
|   |   ├── patterns.json                    # Main database (540 KB)
|   |   └── extraction_cache.json            # LLM extraction cache
|   ├── rag_search_engine.py                 # Search engine implementation
|   ├── pattern_database_builder.py          # Database building pipeline
|   └── drawing_pattern_extractor.py         # LLM-based extraction
├── search_interface.html                    # Interactive search UI
├── serve_search.py                          # Local web server
├── build_comprehensive_sign_index.py        # Sign indexing script
├── drawing-and-anamnestic-signs-features.txt # Reference library (1,305 signs)
├── export-filtered-data-92911-2025921.csv   # Source source document data (654 rows)
└── README.md                                # Usage instructions
```

Dependencies

Python packages:

- `google-generativeai`: LLM API client
- `python-dotenv`: Environment configuration
- `csv`, `json`, `re`: Data processing

Environment variables (`.env`):

```
GEMINI_API_KEY=your_api_key_here
USE_EXTERNAL_LLM=1
```

Performance characteristics

Search performance:

- Pattern search: < 100ms for typical query
- Sign search: < 50ms for keyword matching
- Database load time: < 200ms (540 KB JSON)

Processing performance:

- LLM extraction: ~30-60 seconds per source document
- Batch of 10 source documents: ~5-10 minutes
- Full 654 source documents: ~6-10 hours (estimated)

Database size scaling:

- Current (13 source documents): 540 KB
- Projected (100 source documents): ~2-3 MB
- Projected (654 source documents): ~15-20 MB

Metadata and citation

- **Title**: Rajzvizsgálati adatbázis / Projective Drawing Database
- **Author**: Vass, Zoltan
- **Year/Version**: 2025 / 1.0
- **Publisher**: Károli Gáspár Reformed University (Budapest, Hungary)
- **URL**: https://krepozit.kre.hu/
- **Citation (APA 7)**:
  Vass, Z. (2025). *Rajzvizsgálati adatbázis / Projective Drawing Database* (Version 1.0) [Methodology]. Károli Gáspár Reformed University. Budapest, Hungary. https://krepozit.kre.hu/
- **Collection period**: January 6, 2025 - May 15, 2025
- **Data source**: collected case studies and professional documentation

---

For questions, contributions, or access requests, please contact the project author as listed in the Metadata.

Funding statement

Az adatbázis a Károli Gáspár Református Egyetem Mesterséges Intelligencia Kutatócsoport keretében készült, a "Mixture-of-Experts (MoE) architektúrájú mesterséges intelligencia rendszer készítése egyetemi oktatók publikációs tevékenységének segítéséhez" című, 66020R800 témaszámon támogatott belső projekt vállalásaként, melyet a Károli Gáspár Református Egyetem tudományos projektek támogatására kiírt pályázati konstrukció keretében finanszírozott.

Appendix A: Complete sign category listing

1. ANAMNÉZIS (179 signs)

General anamnestic data, demographics, psychiatric diagnosis, body type, temperament, handedness, emotional spectrum, physical appearance, non-verbal expression, parental family, social behavior, sexual behavior, emotions, cognitive functioning, neurotic symptoms, personality disorders, abuse, addiction, somatic diseases, eating disorders

2. AKTUÁLGENEZIS (114 signs)

Verbal behavior, gestures and facial expressions, test behavior, psychomotor characteristics, drawing time and initial reaction time, drawing sequence, erasing, emotions or thoughts projected into drawing, comparison of drawings made at the same time, other process analysis aspects

3. INTUITÍV ELEMZÉS (6 signs)

Observation with receptive attitude, spontaneous attention focus, motor empathy, kinesthetic empathy, visualization, empathic questioning, single-sentence characterization

4. EGÉSZLEGES ELEMZÉS (189 signs)

Emotional-mood tone, integration and harmony, Hárdi drawing personality levels, integration level, general harmony, enactive/iconic/symbolic levels, movement and form imagery, color usage, Elkisch analysis, style, further holistic aspects

5. FORMAI-SZERKEZETI JELLEMZŐK (89 signs)

Size and proportions, symmetry, position, line quality, shading, detail level, perseveration and repetition, openness-closedness, transparency, distortions and omissions, graphemes

6. TARTALMI-SZIMBOLIKUS JELLEMZŐK (96 signs)

Eleven areas of content-symbolic analysis, including iconic identification, semantic connections, cultural influences, situational factors, personal relationships, associations, communication analysis, manifest and latent content, symbol analysis layers, psychoanalytic analysis, Kapitány motivational analysis

7. HÁZRAJZ (71 signs)

Thematic stereotypes, door, windows, roof and chimney, walls, stairs and pathway, environment and accessory details

8. FARAJZ (83 signs)

Holistic tree characteristics, thematic stereotypes, species, foliage and branches, tree trunk, root and lower trunk, hollow, special markings

9. ÁLLATRAJZ (93 signs)

Animal species (domestic, play/immature, herbivores, predators, insects/spiders, non-existent, birds, mammals, other typical), movement and posture, head and features, body trunk, limbs, latent animal image

10. EMBERRAJZ (299 signs)

Gender, age, thematic stereotypes, posture and movement, head and neck, face and features, body trunk, arms/hands/fingers, legs/feet/toes, hair and body hair, shoulders, breasts, hip and genital area, clothing

and accessories

11. SZABADRAJZ (86 signs)

Thematic stereotypes, other typical representation forms, accessory elements (ground line and background, objects, natural elements, other accessory elements)

Total: 1,305 reference signs across 11 categories (1,116 unique indexed signs in current database)