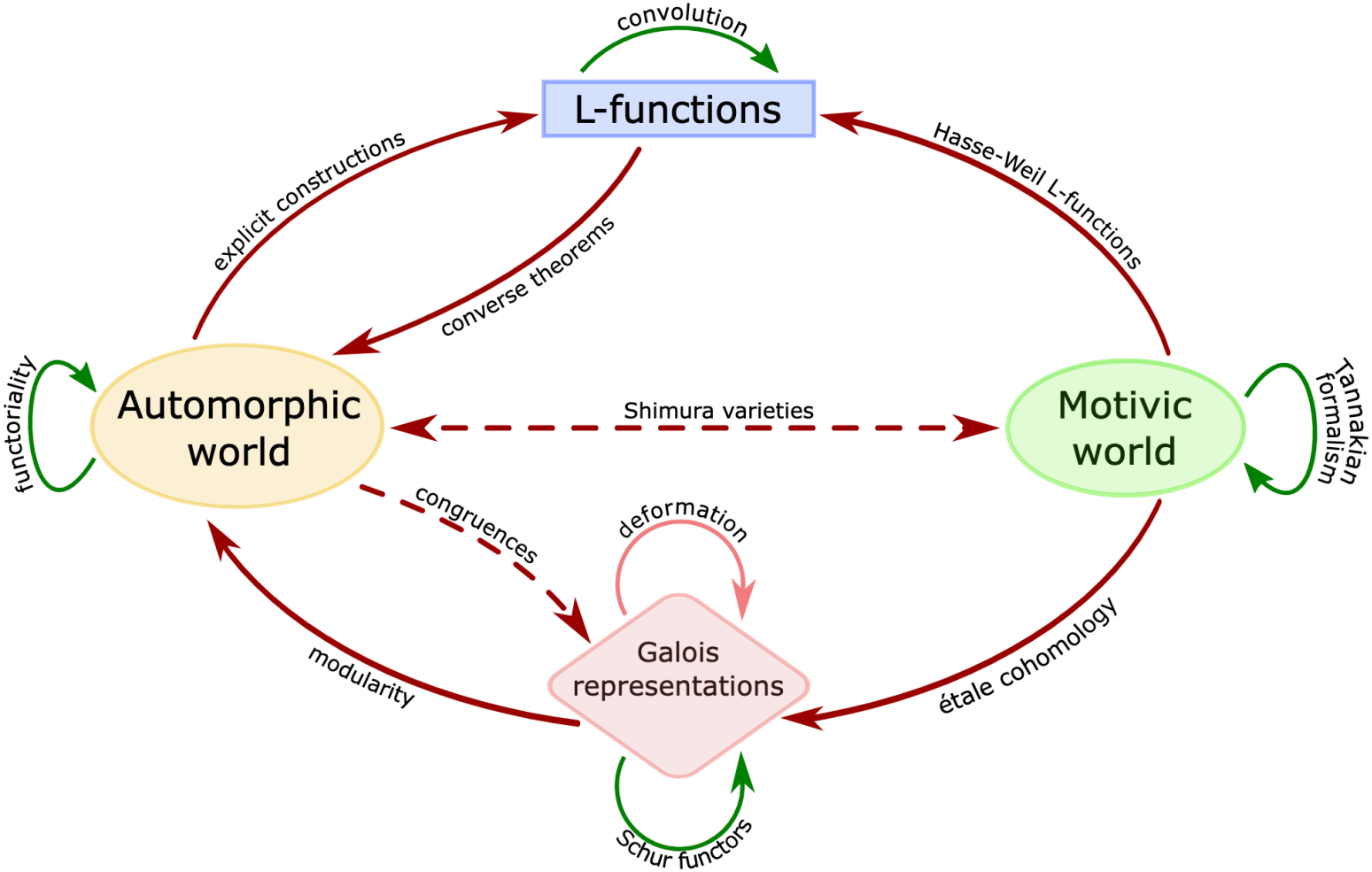
LMFDB: Past, Present and Future

David Roe

Department of Mathematics
MIT

June 29, 2026
Bridging Lean and the LMFDB

The LMFDB is a collection of databases centered around the Langlands program, which relates complex analytic functions (like L-functions and zeta functions), Galois representations (such as Tate-modules of elliptic curves) and automorphic forms and representations (modular forms).



Sections of the LMFDB

- **L-functions** – rational and otherwise
- **Modular forms** – classical, Hilbert, Bianchi, Maass, Siegel
- **Varieties** – elliptic curves over $\mathbb{Q}$ and over number fields, genus 2 curves, modular curves, higher genus families, abelian varieties, Belyi maps
- **Fields** – number fields and p-adic fields
- **Groups** – abstract finite groups, permutation groups, Sato-Tate groups, and lattices
- **Representations** – Artin representations and Dirichlet characters
- **Motives** – hypergeometric

Aims to make deep mathematical connections concrete.

Demo

beta.lmfdb.org

Statistics

Size

About 3TB static (Riemann zeta zeros and class groups of imaginary quadratic fields) and 2TB searchable data, 450GB of search indexes (more details).

Number of objects

24M L-functions, 850k modular forms, 7.5M varieties (with 34M modular curves coming soon), 21M fields, 1M groups, 30M Dirichlet characters, 650k Artin representations, 60k families of hypergeometric motives.

Statistics for each section

Most sections of the LMFDB have a statistics page.

Web interface

- Three versions of the website:
  1. lmfdb.org – hosted on google cloud,
  2. beta – hosted at MIT, includes more preliminary sections like modular curves, hypergeometric motives and lattices.
  3. alpha – hosted at MIT, unfinished sections like Shimura curves and mod-ℓ Galois representations, and in-progress updates to existing sections.
- Runs an open source Flask frontend, allowing for templated webpages that are filled in from the database.
- The backend uses PostgreSQL, with data access also available through the API or Python.

Features

- Knowls – often used for defining terms, expand on the same page. Editable online.
- Aim at both experts and novices. Both search and browse modes, interesting examples.
- Completeness – describe regimes in which data is complete. Search results now show if complete.
- Source and reliability – describe how data was computed, any assumptions used and verifications made
- Results can be further refined, resorted, with displayed columns customized.
- Properties box (often with a picture) and related objects.
- Aim to have downloads in various formats.
- Each object has a mathematically meaningful label and homepage.

Origins

- Idea at a 2007 AIM workshop, work began in 2007; combined existing datasets in number fields, p-adic fields, elliptic curves
- Much early development at workshops, decisions made by consensus
- Transitioned to editorial board in 2018
- From 2018–2025, development driven by Simons Collaboration on Arithmetic Geometry, Number Theory, and Computation
- Current state is strong: weekly development meetings (hybrid), 1250+ citations, many ongoing projects.

Ongoing projects

Recently made ongoing projects visible at alpha.lmfdb.org.

- **Modular forms** – Eisenstein series, updated Siegel modular forms
- **Varieties** – many more genus 2 curves, Shimura curves, expanded modular curves, Hilbert modular surfaces, abelian varieties over finite fields up to isomorphism
- **Representations** – mod-ℓ Galois representations
- **Groups** – updated lattices

AI and mathematical data

- LLMs are getting good at many things: writing code, searching the literature, creating proofs
- Still bound by computational complexity, and people want results quickly. Value remains in **collecting** reusable computations that take a long time.
- Collecting computational results also allows machine-assisted **search** for interesting examples.
- At a technical level, MCP servers (Model Context Protocol) provide ways for LLMs to interact with computational resources.

Formalization

We want to provide users reasons to **trust** data.

Present: informal

- Each section has a reliability knowl
- Different levels of verification; modular forms and modular curves undertook systematic searches for errors
- No uniform standard across different sections

Future: formal

- LLMs make formalization more accessible.
- Some knowls already offer links to Mathlib for definitions.
- Hope to offer Lean downloads for each page, proving a subset of the claims.
- Completeness claims are a challenge.

Questions?

Next up: workshop projects