1. Introduction

This proposal concerns fundamental research in several mathematical areas brought together by the Langlands program. The conjectures in the Langlands program have served as a road-map for research, motivating many of the deepest results in number theory and related areas. The Langlands program can also serve as a guide to navigate between different subfields, thereby providing useful resources to other areas. That is, results in one area have implications for other research areas, via the connections predicted by the Langlands program. Articulating such connections, with the specific goal of developing a better understanding of the impact that objects in one area have on those in another, will be a major outcome of this project. These insights will be in addition to the progress we anticipate making in each area separately.

To facilitate the building of connections between the various subfields, a comprehensive repository of material on many number theoretic objects is in development–and constitutes a unique feature of this project. Our results will be incorporated into the *L-functions and Modular Forms DataBase* (LMFDB), available at `http://www.LMFDB.org/`. The LMFDB will become an ongoing, up-to-date resource for mathematicians and others interested in number theory. The LMFDB organizes a variety of number theoretic objects into individual home pages, showing key data. In many cases, the LMFDB contains complete tables of these objects, such as all elliptic curves over $\mathbb{Q}$ with conductor less than a given bound. Links within the site exhibit the connections between different objects. For example, an elliptic curve over $\mathbb{Q}$, its associated weight 2 cusp form, and their L-functions are all linked. A screen-shot of an elliptic curve homepage is shown in Figure 1.

While most researchers in number theory have a general conception of the connections predicted by the Langlands program, the necessary investment in time inhibits many people from developing an understanding of the areas outside their specialty at the depth required to exploit those connections. Our view is that it is still possible for these researchers to gain an appreciation *for the implications for their own area* without necessarily understanding the underlying details. Conveying these implications more widely can only be beneficial to the field, especially since the new information obtained by transferring knowledge from one area to another can be nontrivial. In Section 1.1 we give several recent examples that have arisen from the LMFDB and its initial development. These examples illustrate how each area can gain new insight by understanding and exploiting the connections to other areas, and that the LMFDB fosters such exchanges.

We will develop the connections that exist between these objects, expressed in terms that are accessible from a variety of perspectives. These will be illustrated by examples that allow people to appreciate the connections without necessarily having to understand all the subtleties of another field. This will be supplemented by enormous collections of data which currently are in various locations in fragmentary form. Bringing together this material will make it possible to make new connections, with the potential for a wealth of new discoveries. This work will require basic research in each area, computational and experimental work to reduce abstract examples to explicit terms, and a sharing of knowledge across disparate fields. The PIs, senior scientists, and LMFDB collaborators on this proposal have expertise in these areas and are committed to working together to develop these connections and to making the results accessible to the wider research community. The result will be an LMFDB that provides an entryway for students and non-specialists, while also leading to new research in all these fields.

Initially, we will develop these connections in terms of L-functions, this being a natural choice because L-functions often serve as the intermediary which links different objects. Later, we will lay the groundwork for a similar treatment of Galois representations, since many objects in the LMFDB give rise to Galois representations. One of the challenges in developing this global picture is the fact that there is little overlap between the experts in the different subfields. Another challenge is that the L-functions connected to each area are viewed differently depending on the source of the L-function. For example:

- From an analytic/computational perspective, an L-function is typically seen as a Dirichlet series satisfying a specific functional equation. The Euler product gives information about the Dirichlet coefficients, but otherwise plays a secondary role. Little use is made of the Diophantine properties of the coefficients.

- For an algebraic variety, the L-function is an Euler product where the local factors contain information about the reduction of the variety mod p. The local factors at the good primes are relatively easy to find, but the bad factors can take considerably more work. The coefficients of the Euler factors have a distribution, dictated by the Sato-Tate group of the curve. The functional equation of the L-function is relatively straightforward to determine, except for the sign and the conductor/level.
- For a Galois representation, the L-function is an Euler product where the factors encode information about the trace of Frobenius. It is not typical to make use of the functional equation, although the level of the representation equals the level of the L-function, and the weight of the representation provides some information about the Γ-factors.
- An automorphic representation can be written as a product of local data over the primes, and the associated L-function arises naturally as an Euler product. The functional equation can be determined explicitly from the representation, although most researchers in the area tend to use the fact that the functional equation exists, without writing it out explicitly.
- For a number field, its Dedekind zeta function has a simple expression as a product over the prime ideals of the field, and its functional equation is straightforward to determine from the discriminant and the number of real and complex embeddings. The zeta function factors into a product of Dirichlet L-functions and Artin L-functions.
- For modular/automorphic forms in one variable the L-function is obtained via a Mellin transform. But the situation is more complicated in general. For example, for Siegel and Hilbert modular forms the Hecke eigenvalues, which are used to define the L-function as an Euler product, are not the Fourier coefficients. At present it it not known how to reconstruct a Siegel modular form from its L-function.

The above examples illustrate that L-functions have a wide variety of properties, but most areas use only a subset. This means that in most cases information is available but is not being used. That progress can be made by utilizing this information is reinforced by the following examples.

1.1. **Some recent connections.** The following list shows some of the transfer of information between fields which has occurred during the early development of the LMFDB. We also give an example of how collecting data in a central location and providing new ways to visualize it can lead to new discoveries. These are fairly modest examples, but they indicate the potential of the LMFDB to support research across areas in which communication currently is limited.

(1) *Automorphic forms → classical L-functions*: Calculation of the Γ-factors of Maass form L-functions revealed a misconception about the possible Γ-factors in the degree 2 case, and revealed a gap in the literature in the degree 3 case. See Section 2.1.
(2) *Classical L-functions → automorphic forms*: The strong multiplicity one theorem for L-functions is both stronger and easier to prove than the analogous statement for automorphic representations. It implies that if two Siegel cusp forms on $Sp(4)$ have the same eigenvalues for the Hecke operators $T(p)$, then they have the same eigenvalues for $T(p^2)$. See [67].
(3) *Varieties ↔ classical L-functions*: Precise knowledge of the possible bad factors of a hyperelliptic curve made it possible to directly search for such L-functions, which in turn allows one to prove the non-existence of such a curve with a given conductor. See Section 3.3.
(4) *Classical L-functions → elliptic curves*: The plot of the L-function of the first rank 4 elliptic curve (see the homepage on the LMFDB) shows unusual behavior near $t = 14$, 21, and 25. Michael Rubinstein explains this by the elliptic curve L-function mimicking $1/\zeta(1+it)^4$.
(5) *Visualizing data*: For every level up to $N = 11$, the smallest k for which the space of classical modular forms $S_k(\Gamma_0(N))$ is nonempty, the Fourier coefficient a_2 is negative. The same holds for low-level Siegel modular forms on $Sp(4)$. This observation, which does not appear to have been made before, is easy to see if one can visualize the data. Having made the observation, the cause can be explained using the explicit formula [65].

The results of this project, once incorporated into the LMFDB, will vastly increase the variety of objects, the number of connections, and the amount of data available. We expect a proportionate increase in the number and quality of results which are fostered by the LMFDB.

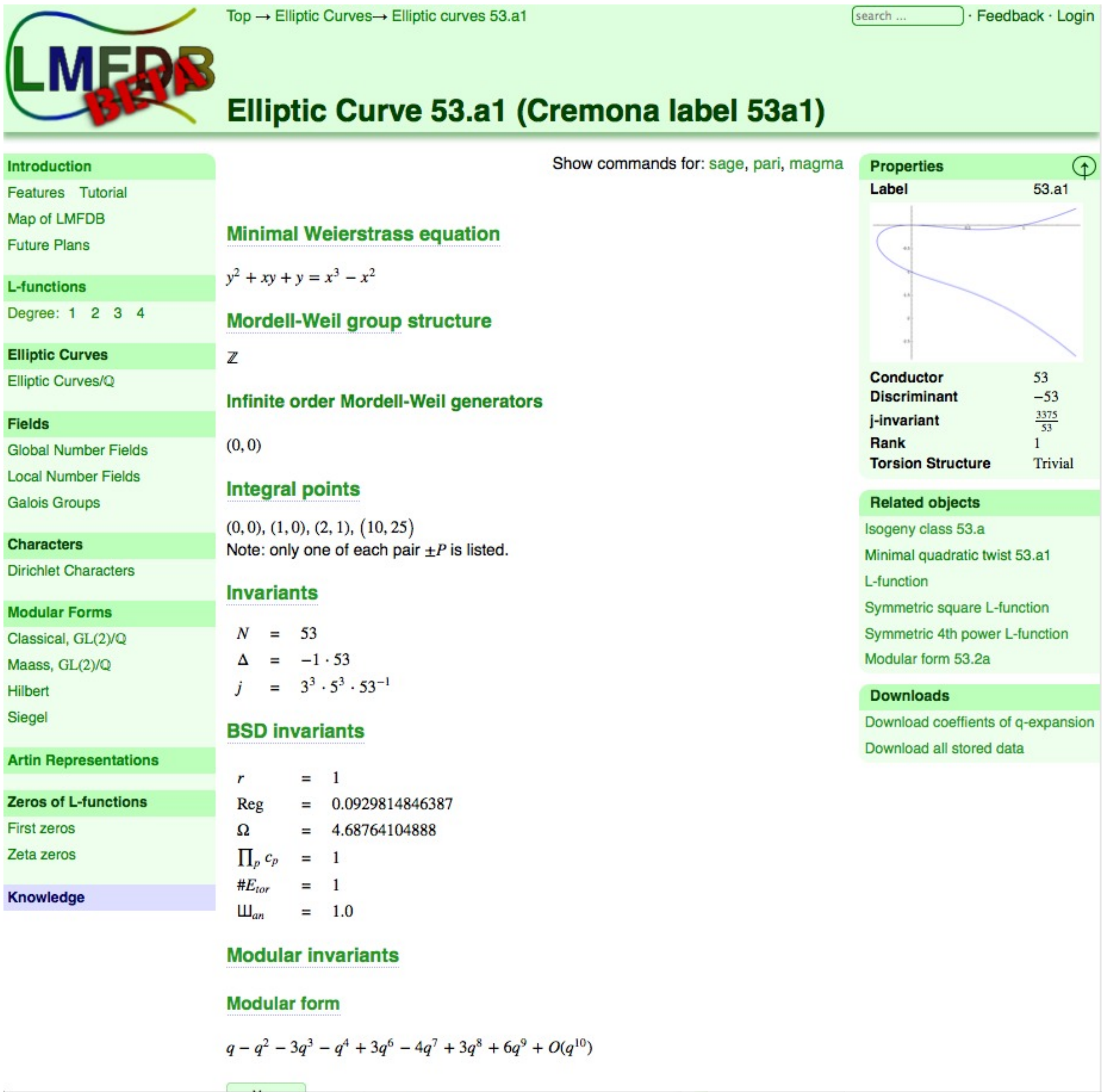

FIGURE 1. The home page of an elliptic curve. Note that the page has a comprehensive description of the properties of the elliptic curve, links to more data, links to sage, pari, and magma commands which generate the data, a picture, and links to related objects, such as the associated L-function and the associated modular form.

1.2. **Our approach.** In this introduction we have focused on the LMFDB as a resource for the research community. While this is the most visible outcome of our proposed work, the vast majority of the activity described in this proposal is a combination of traditional research in the areas covered by the LMFDB along with research which leads to the creation of data which are useful for forming and testing conjectures.

At twice-yearly workshops the researchers on this project, along with members of the broader mathematics community, will come together to incorporate their results into the LMFDB. These workshops will serve to bring new people into the project, in addition to the obvious benefit of making the LMFDB an increasingly comprehensive resource.

2. L-FUNCTIONS AS AN ORGANIZING PRINCIPLE

One view of L-functions is that they unite different objects: number fields, varieties, automorphic representations, modular forms, and Galois representations all have L-functions associated to them. From this point of view, L-functions serve as an organizing principle. The data in the functional equation of the L-function – degree, level, spectral parameters, etc., (see Section 2.1 for details) – imposes a hierarchy and partitions the objects into natural equivalence classes.

Currently the LMFDB organizes each class of objects separately. For example, number fields or elliptic curves each has its own browse and search screen. This, of course, is useful when one is interested in a particular class of objects, or finding an object with particular properties.

We will create a more global perspective by providing a view of the entire subject in terms of L-functions. This will provide a uniform way to navigate the website, and also better emphasize the connections between objects. The first step is to understand the classification of objects in terms of their L-functions, which we describe in the next section.

Note: we will also begin work that will lead to a global view of (most of) the LMFDB in terms of Galois representations. This is described in Section 11.

2.1. **Classification of L-functions: functional equation.** A fundamental question is: What are the possible functional equations of an L-function?

An answer is provided by representation theory: the L-function of an automorphic representation satisfies a functional equation, and a calculation with the Weil-Deligne group determines the precise data of the Γ-factors, level, and sign. Unfortunately, there seems to be very little overlap between the people with the automorphic perspective (who generally don't need to specify the functional equation in concrete terms) and the people with an analytic or computational perspective (who would like to know every detail of the functional equation). For example, Borel's account in the Corvallis proceedings [30] is not easily accessible without significant background knowledge.

In work which will be nearing completion at the start of this proposed project, PI Farmer, senior scientist Schmidt, and collaborators Pitale and Ryan, have been recasting the automorphic perspective, in particular the calculations with the Weil-Deligne group, in terms that are accessible from an analytic perspective. This work provides a good example of the benefits in bringing together areas which traditionally do not have close communication.

The functional equation of an automorphic L-function $L(s) = \sum a(n)n^{-s}$ can be written as

$$\Lambda(s) := N^{s/2} \prod_{j=1}^{J} \Gamma_{\mathbb{R}}(s+\mu_j) \prod_{k=1}^{K} \Gamma_{\mathbb{C}}(s+\nu_k) \cdot L(s) = \varepsilon \overline{\Lambda}(1-s), \tag{2.1}$$

where the positive integer N is called the *level*, and the *spectral parameters* satisfy $\mathrm{Re}(\mu_j) = 0$ or 1, and $\mathrm{Re}(\nu_k)$ is a positive integer or half-integer. Here $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$ and $\Gamma_{\mathbb{C}}(s) = (2\pi)^{-s}\Gamma(s)$. Note that (2.1) is written in the so-called analytic normalization which relates s to $1-s$.

A feature of (2.1) is that it is unique: there is only one way to write the functional equation in that form. This means that the data in the functional equation, $(d, N, (\mu_1, \ldots, \mu_J : \nu_1, \ldots, \nu_K), \varepsilon)$ can be used to classify L-functions and the objects giving rise to them. Here $d = J + 2K$ is the *degree* of the L-function, which is the crudest measure of its complexity. The Riemann zeta function and Dirichlet L-functions have degree 1, L-functions of classical holomorphic cusp forms and Maass forms have degree 2, and so on.

Another benefit of making these representation theoretic calculations available to the classical viewpoint is that the relationship between L-function data and features of the underlying object can be better understood. For example, it is well known that for $\mathrm{SL}(2,\mathbb{Z})$ Maass forms, the L-functions come in two forms, with Γ-factors

$$\Gamma_{\mathbb{R}}(s+ir)\Gamma_{\mathbb{R}}(s-ir) \qquad \text{or} \qquad \Gamma_{\mathbb{R}}(1+s+ir)\Gamma_{\mathbb{R}}(1+s-ir). \tag{2.2}$$

Those two cases correspond to even and odd Maass forms, respectively. What is not widely known is that for Maass forms on $\Gamma_0(N)$ with character χ, the same Γ-factors occur *provided* χ *is even.* If

χ is an odd character, then the Γ-factors are $\Gamma_{\mathbb{R}}(s+ir)\Gamma_{\mathbb{R}}(1+s-ir)$. As far as we know, nobody has ever actually done a numerical calculation (checking the Riemann Hypothesis, for example), for such an L-function. The need for this foundational work is underscored by the fact that even this degree 2 case was not generally understood.

For Maass forms on $\mathrm{SL}(3,\mathbb{Z})$, the L-functions should again come in two forms, with Γ-factors

$$\Gamma_{\mathbb{R}}(s+ir_1)\Gamma_{\mathbb{R}}(s+ir_2)\Gamma_{\mathbb{R}}(s+ir_3) \qquad \text{or} \qquad \Gamma_{\mathbb{R}}(1+s+ir_1)\Gamma_{\mathbb{R}}(1+s+ir_2)\Gamma_{\mathbb{R}}(s+ir_3), \tag{2.3}$$

where the r_j are real and $r_1+r_2+r_3=0$. However, when these functions are treated in classical terms in the literature, such as the books of Bump [38] or Goldfeld [77], only the first set of Γ-factors apears. Bump is explicit about working over $\mathrm{GL}(3,\mathbb{Z})$ and on page 70 of his book uses a matrix of determinant -1 to (essentially) restrict to the analogue of the even case of $\mathrm{SL}(2,\mathbb{Z})$. Goldfeld works over $\mathrm{SL}(3,\mathbb{Z})$, so presumably there is a second case, coming from the fact that the integrals on page 183 of his book could vanish identically. We will resolve this apparent discrepancy. This is an appropriate project for a student.

A specific goal of this project is to work out, in completely explicit terms, all possible functional equations up through degree 4, and to match these with all known sources of L-functions. The odd case of $\mathrm{SL}(3,\mathbb{Z})$ is a good project for a student, because it brings together both perspectives on L-functions in a context where the general outline has already been set out. We expect to identify many other specific cases of interest that would also be suitable as student projects.

In parallel to this we will use the methods described in Section 10.1 to directly search for all possible L-functions up through degree 4, having parameters within a reasonable bound. We expect to complete the case of (2.3) very early in this project, but other cases, particularly in degree 4, will require considerable work.

A more lengthy endeavor, which will see steady progress throughout the project, is to implement a global organization to the LMFDB based on L-function data. This involves more information than just the functional equation, which we discuss in the next section.

2.2. **Classification of L-functions: coefficients.** The functional equation of an L-function only goes part way toward a complete classification. An important subset is those L-functions which are *algebraic*, meaning that (suitably normalized) they have algebraic integer Dirichlet coefficients. Conjecturally, these are the L-functions for which, in the notation of (2.1), μ_j and ν_k are all real. Again conjecturally, these are the L-functions which are motivic, with motivic weight given by $w = \max\{0, \nu_1, \ldots, \nu_k\}$. This means that $b(n) = a(n)n^{w/2}$ is an algebraic integer, and the arithmetically normalized L-function $L_{ar}(s) = \sum b(n)n^{-s}$ satisfies a functional equation relating s to $w+1-s$. Note that motivic weight is different than the classical weight for GL(2) cusp forms, and it is different yet again to the classical weight of Siegel modular forms.

A current shortcoming of the LMFDB is that L-funtions are only shown in the analytic normalization. This will be remedied in Workshop 1 of this project. That work will be led by PI Farmer and senior scientist Rubinstein, and will be done in parallel with the addition to the LMFDB of hyperelliptic curves, led by PI Kedlaya and senior scientist Sutherland. This will allow more researchers to view L-functions in a way which is more natural to them, and will serve as preparation for adding Galois representations to the LMFDB (see Section 11).

There are various other properties of L-functions which could form the basis for organizing the LMFDB: self-duality (i.e., real coefficients), rank (order of vanishing at the critical point), primitivity (is it a product of other L-functions), lift or twist (does it arise from another L-function of smaller degree or level), etc. All L-functions will be assigned these properties, and these properties will be used for searching and browsing the site. But there is one additional property which we believe is of key importance: the distribution of coefficients.

Conjecturally, every L-function has a "Sato-Tate group" associated to it. See Serre [123] for a detailed treatment. The Sato-Tate group of L is a group $G_L \subset U(d)$ of $d \times d$ unitary matrices, where d is the degree of the L-function. The conjecture is that the eigenvalues of G_L have the same distribution, under Haar measure on $U(d)$ restricted to G_L, as the distribution of the Satake parameters of the L-function. Here the Satake parameters are defined in terms of the Euler product:

$$L(s) = \prod_p F_p(p^{-s})^{-1}, \tag{2.4}$$

where F_p is a polynomial: $F_p(z) = (1 - \alpha_{1,p}z) \cdots (1 - \alpha_{d,p}z)$. If p is a good prime (i.e., p does not divide the level) then the $\alpha_{j,p}$ are the Satake parameters. In the analytic normalization, the Ramanujan conjecture says $|\alpha_{j,p}| = 1$ at the good primes.

PI Kedlaya and senior scientist Sutherland (see Section 3.1) have classified the Sato-Tate groups for hyperelliptic curves of genus 2 and will be addressing the genus 3 case. We will introduce the Sato-Tate groups as invariants of objects in the LMFDB, and use them as tools for studying L-functions. PI Farmer made some preliminary calculations in the hypergeometric case (Section 9) which suggests that there are two commonly occurring distributions for those L-functions, with the other distributions occurring rarely. The Sato-Tate groups will also become an important input to the operations on L-functions (Section 8.2), because these distributions determine, for example, which symmetric powers of the L-function will have poles.

3. Algebraic varieties

Algebraic varieties provide an essential source of L-functions. The LMFDB currently contains data for varieties in dimension 0, i.e., number fields, and for a class in dimension 1, elliptic curves over $\mathbb{Q}$. We propose to expand to larger classes of varieties.

Our initial efforts will focus on hyperelliptic curves/$\mathbb{Q}$ (Section 3.1) and elliptic curves over quadratic fields (Section 4.1). These have degree 4 L-functions, which makes them a key part of our goal to develop a complete picture of low-degree objects. These also provide important (and large) classes of Galois representations, laying the groundwork for our development of Galois representations as both an organizing principle and as basic objects in the LMFDB (Section 11).

In the later part of the project we further develop elliptic curves over quadratic and cubic fields (Sections 4.1-4.2), and we will begin considering other varieties, particularly those which are expected to be modular (Section 3.2).

3.1. **Hyperelliptic curves.** Among the class of algebraic curves, hyperelliptic curves on one hand form the most easily studied subclass due to the explicit nature of their description, and on the other hand are general enough to exhibit many phenomena typical of arbitrary algebraic curves.

Achieving a meaningful and useful classification of hyperelliptic curves in the LMFDB, comparable to the Cremona database of elliptic curves, would require progress in the following directions. (Throughout this discussion, we will limit attention to hyperelliptic curves over $\mathbb{Q}$, but we may consider working over other number fields later.)

(1) *Tabulation of Euler factors at primes of good reduction.* For genus 2 and 3, this is largely addressed by the work of Kedlaya and Sutherland [108], [132], which incorporates p-adic cohomology in the form of an algorithm of Harvey [90] along with a highly optimized baby-step giant-step algorithm for smaller primes.

(2) *Computation of Euler factors and other associated data at primes of bad reduction, e.g., the conductor.* This is most tractable in genus 2, where an implementation of an algorithm due to Qing Liu (which computes the conductor and reduction type for odd residue characteristic) is available in Sage. Extending this to the case of residue characteristic 2 is not straightforward, but an algorithm for computing regular models is implemented in Magma and another one is currently being implemented in Singular. Using this, one can find the Euler factors using the Galois action on the dual graphs of the special fibers of the regular model and the Euler factors for components of positive genus in those special fibers. To find the conductor one needs a semistable model. Algorithms for computing this and deducing the conductor are currently being developed by Wewers. Alternatively, one can conditionally collect some of the same data using analytic methods, e.g., methods of Booker [29] or generic methods using the approximate functional equation (see Section 8.1).

(3) *Computation of Mordell-Weil groups of Jacobians of hyperelliptic curves.* Computing generators of a subgroup of finite index may be facilitated by the recent work of Bhargava and Gross on 2-Selmer groups of hyperelliptic curves [26], but some work (suitable for a graduate student or postdoc) is still needed. In order to compute generators for the full Mordell-Weil group, one needs to saturate a given finite index subgroup. For $g = 2$, everything has been implemented in Magma; for $g = 3$, this is addressed by work in progress of Stoll. For arbitrary g, canonical heights can be computed using the algorithm from [115] (implemented in Magma). In addition, one needs to bound the difference between canonical and naive heights and search for points of naive height up to a given bound. Recent work of Holmes [94] provides such an algorithm, but to make it practical the resulting bounds will have to be improved dramatically, probably using techniques from Arakelov geometry.

(4) *Computation of other BSD invariants, e.g., torsion subgroups, regulators, and Tamagawa numbers.* In genus 2, many of these questions are addressed in existing literature (see for instance [40]) but are not fully implemented. The computation of torsion subgroups is implemented in Magma. To compute Tamagawa numbers, one needs the configuration of the special fibers of a regular model (see (2) above) and how Galois acts on the components. Given a finite index subgroup, one can compute the regulator (up to a rational square) using algorithms due to Müller [115] (implemented in Magma) or Holmes [95]. For $g = 2$ (and soon for $g = 3$ due to work in progress of Stoll) one can compute the regulator exactly using Magma. See (2) above.

(5) *Computation of p-adic regulators and comparison with special values of p-adic L-functions.* Balakrishnan and Besser [18] describe a method to compute p-adic heights via Coleman integration. Putting this together with work of Müller [115] gives us a means of computing p-adic regulators. Pollack and Stevens [120] gave an algorithm that computes special values of p-adic L-functions to high precision using overconvergent modular symbols. Balakrishnan, Müller, and Stein [116] recently stated a p-adic analogue of the Birch and Swinnerton-Dyer conjecture for modular abelian varieties and numerically compared p-adic special values to p-adic regulators, thus giving evidence for the conjecture. PI Stein's Ph.D. student Simon Spicer is doing his thesis work on optimization and applications of the overconvergent modular symbols algorithm, which would be partly supported by this grant.

(6) *Determination of all curves of a given genus with conductor in a given range, assuming appropriate modularity conjectures.* This is probably only feasible for genus 2, and in any case is closely related to Siegel modular forms (see Section 6). For small conductors we can make use of L-function techniques (see Section 3.3).

(7) *Rational points.* The most promising methods for this are variants of the Chabauty method, possibly combined with the Mordell-Weil sieve, see for instance [36]. Both of these depend on (partial) knowledge of the Mordell-Weil group. In genus 2, this should cover most cases encountered in practice; see for example [35].

(8) *Sato-Tate distributions (or equivalently motivic Galois groups).* These are fully classified in genus 2; see [71]. We will consider genus 3 next.

3.2. **Other varieties.** In the long run the LMFDB will include other varieties, initially focusing of those which are modular. We see the proposed work on hyperelliptic curves and elliptic curves over number fields as providing the foundation for how the LMFDB will handle varieties and their relationship with other objects. We anticipate some initial testing of more exotic examples, so that it will be easier to introduce more complicated objects later.

3.3. **Hyperelliptic curves of small conductor.** At present there exist extensive tables of elliptic curves/$\mathbb{Q}$ which are both accurate and complete [52]. These tables are made possible by three theoretical results: the fact that elliptic curves are modular [32], the fact that there exists a method (modular symbols) for finding a basis of the space of holomorphic cusp forms of a given weight and level, and also the fact that one can directly construct the elliptic curve from the modular form.

For hyperelliptic curves, none of those results are currently known. In fact, it was only recently [37] that the proper modularity conjecture was even formulated, indicating precisely which weight 2 Siegel modular forms should be associated to hyperelliptic curves of genus 2.

Developing new ways to generate Siegel modular forms is a major undertaking of this proposal (see Section 6). But before such results are available, there is another approach, using L-functions, which we can use to develop a provably complete table of hyperelliptic curves of small conductor.

Some progress in this direction is already known:

Theorem[Brumer and Kramer] *Suppose A is a semistable abelian surface of odd non-square conductor N. If $N \le 500$, then N can only be 249, 277, 295, 349, 353, 389, 427, 461, for which examples are known, or 415, 417, which we expect do not occur.*

Thus, at least for odd conductors, the list of known hyperelliptic curves is complete, at least up to 413. The reason that Brumer and Kramer cannot resolve the case of $N = 415$ is that their method for eliminating a conductor (for which they suspect there is no corresponding hyperelliptic curve) is to show the nonexistence of a number field with certain properties. But for $N = 415$ there does exist such a field, but presumably no corresponding hyperelliptic curve, so their method fails.

We propose using analytic L-function techniques to handle some cases which are not accessible by algebraic methods, leading to a provably complete table of hyperelliptic curves of small conductor. This is an illustration of our theme that new results follow from bringing together different areas.

A genus 2 hyperelliptic curve has a degree 4 L-function which (in the analytic normalization) satisfies a functional equation of the form

$$\Lambda(s) = N^{s/2}\Gamma(s+\tfrac{1}{2})^2 L(s) = \pm\Lambda(1-s), \tag{3.1}$$

where N is the conductor of the curve. In that normalization, the Dirichlet coefficients are of the form $a_p = A(p)/\sqrt{p}$ where $A(p)$ is a rational integer with $|A(p)| \le 4\sqrt{p}$.

Work of PI Farmer, together with Koutsoliotas and Lemurell, shows

Proposition[FKL] *There is no L-function satisfying* (3.1) *with $N = 415$, with Dirichlet coefficients $a(p^j) = A(p^j)/p^{j/2}$ where $A(p^j)$ are rational integers satisfying $|A(p)| \le 4\sqrt{p}$ and $|A(p^2)| \le 10p$.*

Thus, assuming a modularity conjecture, there is no hyperelliptic curve of conductor 415.

The idea behind the calculation is that one can use the approximate functional equation to write down equations for the Dirichlet coefficients, as in Section 10.1. But instead of solving the system, one checks the consistency of the system using the (finitely many) possibilities for the first few coefficients and the Hasse bound for the remaining coefficients. This eliminates some of the options for the first few coefficients. Then one proceeds to search the tree of possibilities for the later coefficients, pruning branches when the initial choices are incompatible with the Hasse bound for the later coefficients. If such an L-function exists, then this method finds its first several Dirichlet coefficients. If no such L-function exists, then one can hope that every branch of the tree is pruned and the non-existence becomes a theorem.

The above proposition is work-in-progress, but based on our experience in that calculation, we expect to determine the complete list of hyperelliptic curve L-functions up to conductor 1000, and probably a bit further. This would bring the table of hyperelliptic curves up to the point where elliptic curve tables were around 1990. Considering the enormous research benefits (both direct and indirect) that have come from making detailed tables of elliptic curves, we believe that producing such tables and making the data public is an important research contribution.

4. Elliptic curves over number fields

In 2012, J. Cremona successfully enumerated all elliptic curves over $\mathbb{Q}$, ordered by conductor, up to the first curve of rank 4 (which has conductor 234,446). This major milestone was the culmination of over two decades of progress, including improvements in theory and implementations of algorithms and access to more powerful computing equipment. Inspired by Cremona's work, we propose to carry out similar projects over number fields other than $\mathbb{Q}$. Work in this section will be coordinated by PIs Gunnells and Stein, with support from senior scientists Ash and Cremona.

Our initial work, described in Section 4.1, will focus on quadratic fields, and more specifically on real quadratic fields. Focusing on totally real number fields F grants us substantial extra structure. In particular, there is a *conjectural* bijection between isogeny classes of elliptic curves over F and rational Hilbert modular newforms of parallel weight 2 (this bijection is only known when $F = \mathbb{Q}$ – see [32]), and this bijection provides functional equations for L-functions, Heegner points, and results toward the Birch and Swinnerton-Dyer conjecture (see [136]).

One missing ingredient is an efficient algorithm to construct the (conjectured) elliptic curve from the Hilbert modular form. We are forced to combine a variety of strategies, described below, which in the course of this project we hope to make more efficient.

We also consider elliptic curves over nonreal cubic fields in Section 4.2. For that case, the underlying theory is less developed, and there is a significant amount of basic research to be done.

4.1. **Elliptic curves over $\mathbb{Q}(\sqrt{5})$.** Our goal is to find all modular elliptic curves over $\mathbb{Q}(\sqrt{5})$, ordered by conductor, up to the first of analytic rank 4.

Noam Elkies found the smallest known conductor of a elliptic curve of rank 4 over $\mathbb{Q}(\sqrt{5})$ that is not a base change from $\mathbb{Q}$; it has norm conductor 1,209,079. We estimate that it should be possible to find all the rational Hilbert modular newforms of norm conductor up to 1,209,079 in around 200,000 hours of CPU time, e.g., less than two months using the hardware we are requesting as part of this proposal. It is difficult to estimate the time it will take to find the corresponding elliptic curves, since there is no known efficient algorithm to find an elliptic curve attached to a newform (in some cases, we do not even know there is a curve, though this is conjectured).

We use an optimized version of the algorithm in [53] to compute all rational Hilbert modular cusp forms over $\mathbb{Q}(\sqrt{5})$ of weight $(2, 2)$ and level $\mathfrak{n}$. In order to implement this algorithm, it is critical that we can compute with $\mathbb{P}^1(R/\mathfrak{n})$ very, very quickly; as explained in [28, §2.2] we have a highly optimized implementation that we have customized for the case $\mathbb{Q}(\sqrt{5})$. Supported by this grant, PI Stein's graduate student R. A. Ohana intends to generalize this approach to other fields.

Some techniques for finding a curve attached a newforms f with Hecke eigenvalues $a_{\mathfrak{p}}(f)$, for primes $\mathfrak{p}$ of the ring of integers of $\mathbb{Q}(\sqrt{5})$:

(1) *Stein-Watkins search* – consult a database made by doing a search in the style of [129].
(2) *Sieved enumeration* – use $a_{\mathfrak{p}}(f)$ to impose congruence conditions on the coefficients of the Weierstrass equation, then search.
(3) *Torsion families* – use $a_{\mathfrak{p}}(f)$ to determine whether $\ell \mid \#E(F)$ for some E attached to f, and if so search for E in the family of curves with ℓ-torsion.
(4) *Congruence families* – if we know some E' and that $E'[\ell] \approx E[\ell]$, use Tom Fisher's explicit families [70].
(5) *Twisting* – find a minimal conductor twist f^{χ} of f, find a curve attached to f^{χ}, then twist it to get a curve attached to f.
(6) *Cremona-Lingham* – find many curves with good reduction outside the level $\mathfrak{n}$ of f by searching for integral points on auxiliary curves [51].
(7) *Dembele, Bober* – determine periods from special values of L-series [54].

Once we find a curve in the isogeny class corresponding to the rational Hilbert newform, Billerey's algorithm [27] combined with Velu's equations for isogenies enable us to explicitly compute representative elliptic curves for each isomorphism class.

In [28], the PI Stein and numerous undergraduates, graduate students, and a postdoc, made a complete table of every elliptic curve over $\mathbb{Q}(\sqrt{5})$, up to the first curve of rank 2 (which has norm conductor 1831). Stein intends to partly support one of these graduate students using this grant (A. Deines) to carry out a Stein-Watkins [129] style search for curves over $\mathbb{Q}(\sqrt{5})$. Another one of the co-authors (R.A. Ohana) developed the sieved enumeration method, and is starting graduate school now at UW; Stein intends to support Ohana to do further work in this direction.

4.2. **Elliptic curves over complex cubic fields.** In this section we describe the computation of tables of elliptic curves over complex cubic fields F (i.e., cubic fields with exactly one real place). Such computations have already been carried out for F the cubic field of discriminant -23 in work of PI Gunnells with Yasaki and Klages-Mundt [82, 87] up to conductors of norm 911. Our goals are to extend these computations in several directions. See Section 5 for background.

As in Section 4.1, one expects that for every elliptic curve E over F, there should exist a cuspform f on GL_2/F with rational Hecke eigenvalues such that $L(s, f) = L(s, E)$. Hence the first step is computing a table of suitable automorphic forms on GL_2/F, so that one has a (conjectural) list of conductors of possible curves. Note that the connections between such forms and arithmetic are completely unproven: one doesn't even know, for example, that such forms will have ℓ-adic families of Galois representations attached to them, in contrast to the Hilbert modular case.

To compute the relevant automorphic forms, we use the fact that they are cohomological and compute the cohomology of subgroups of $\mathrm{GL}_2(\mathscr{O})$, where $\mathscr{O}$ is the ring of integers of F. These methods are quite different than those described in Section 4.1 and we refer to Section 5 for background. For any ideal $\mathfrak{n} \subset \mathscr{O}$, we consider the subgroup $\Gamma_0(\mathfrak{n})$ of matrices upper-triangular mod $\mathfrak{n}$ and form the locally symmetric space $Y_\mathfrak{n} = \Gamma_0(\mathfrak{n})\backslash G/KA_G$. The space $Y_\mathfrak{n}$ has dimension 6, the virtual cohomological dimension ν is 5, and the top of cuspidal range is 4. There is an explicit reduction theory due to Koecher [110], generalizing Voronoi's work on perfect quadratic forms [135], that allows us to construct a cell decomposition of $Y_\mathfrak{n}$ for any $\mathfrak{n}$. Using this we compute $H^4(Y_\mathfrak{n}; \mathbb{C})$ for all ideals up to some bound on their norm. Note that work of Ash [4] provides an analogue of the classical theory of modular symbols for $\mathrm{GL}_2/\mathbb{Q}$ [114] for this setting, but it computes H^5, not H^4, and thus does not see the cuspidal cohomology.

To identify the forms corresponding to elliptic curves, we must compute the Hecke action on H^4. This requires some effort, since even though the Hecke operators act on cohomology, they don't preserve the Koecher cells, and thus do not act directly on the cochain complex built from these cells. The classical theory of modular symbols encounters the same problem. More precisely, consider the Satake compactification $\mathfrak{H}^*$ of the upper halfplane, suppose $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ is a congruence subgroup, and let $X_\Gamma = \Gamma\backslash\mathfrak{H}^*$. Any two cusps u, v determine a modular symbol $[u, v] \in H_1(X_\Gamma, \partial X_\Gamma; \mathbb{C})$: one takes the class of the image of the oriented ideal geodesic in $\mathfrak{H}^*$ running from u to v. The analogues of the Koecher cells are the $\mathrm{SL}_2(\mathbb{Z})$-translates of the geodesic from 0 to ∞; the resulting modular symbols are called unimodular symbols. One can determine the relations among the modular symbols to determine a combinatorial model for the relative homology, which gives a model for $H^1(X_\Gamma \smallsetminus \partial X_\Gamma; \mathbb{C})$ by duality. The modular symbols admit an action of the Hecke operators compatible with duality, but the action does not preserve the subspace of unimodular symbols.

In the setting of GL_2/F, the Koecher cochain complex is the analogue of the space of unimodular symbols, and to compute the Hecke action we embed it in a larger complex, the *sharbly complex*. This complex, which gives a resolution of the Steinberg module of $\mathrm{GL}_n(F)$, is essentially the chain complex of the free simplicial complex on the cusps of a certain compactification of G/KA_G [14, 86, 87]. Computing the Hecke action involves taking a sharbly cycle representing a class in $H^4(Y_\mathfrak{n}; \mathbb{C})$ and rewriting it in terms of cycles supported on Koecher chains. This is similar to the approach used in the context of $\mathrm{SL}_4(\mathbb{Z})$ [10–14, 79], another situation in which the top of the cuspidal range is $\nu - 1$. This relies heavily on a generalization of the Ash–Rudolph algorithm [16] for modular symbols on $\mathrm{GL}_n(\mathbb{Z})$. Computing Hecke operators is the main bottleneck, although one can often predict the existence of elliptic curves at higher level norms by computing ranks of cohomology groups and using a yoga of "old and new" cohomology classes [83].

Once one has a list of possible conductors, one must find equations of elliptic curves in each isogeny class. To do this we use similar techniques as in Section 4.1: naive enumeration, Cremona–Lingham, searching in torsion families, and so on. Stein and Gunnells will coordinate on this work. Once a curve in each isogeny class is found, work of Billerey [27] enables us to compute representatives for each of the isomorphism classes, just as for the $\mathbb{Q}(\sqrt{5})$ case.

We propose to work towards the following goals:

(1) Extension of the tables in [82, 87] as far as possible.
(2) Extending the computations in [82, 87] for other complex cubic fields. Data for elliptic curves over the fields of discriminants $-31, \ldots, -107$ can be found at [109] and in [83].
(3) Extend the computations of [87] to nontrivial coefficients, i.e. compile tables of higher weight cuspforms.
(4) Incorporate all this data into the LMFDB.

5. Cohomology of arithmetic groups and automorphic forms

Elsewhere in this proposal we describe several problems that involve the computational investigation of automorphic forms via the cohomology of arithmetic groups. For reference we collect together some background here. The relevant projects are described in sections 4.2, 6, and 12.

Let $\mathbf{G}$ be a reductive algebraic group defined over $\mathbb{Q}$, and let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be an arithmetic subgroup. Let $\mathscr{M}$ be a $\mathbb{Z}\Gamma$-module. Then the cohomology $H^*(\Gamma; \mathscr{M})$ provides a link between number theory and topology. Specifically, suppose $\mathscr{M}$ arises from a rational representation of $G = \mathbf{G}(\mathbb{R})$ on a finite dimensional complex vector space. Then J. Franke has shown [72] that there is a decomposition

$$H^*(\Gamma; \mathscr{M}) = H^*_{\mathrm{cusp}}(\Gamma; \mathscr{M}) \oplus \bigoplus_{\{\mathbf{P}\}} H^*_{\{\mathbf{P}\}}(\Gamma; \mathscr{M}), \tag{5.1}$$

where the sum is taken over the set of classes of associate proper $\mathbb{Q}$-parabolic subgroups of $\mathbf{G}$. The summand $H^*_{\mathrm{cusp}}(\Gamma; \mathscr{M})$ is the *cuspidal cohomology*; this is the subspace of classes represented by cuspidal automorphic forms. The remaining summands constitute the *Eisenstein cohomology* of Γ. In particular the summand indexed by $\{\mathbf{P}\}$ is built of Eisenstein series (and their residues) attached to certain cuspidal automorphic forms on the Levi factors of elements of $\{\mathbf{P}\}$.

Hence there is a close connection between the cohomology of Γ and automorphic forms. Since the former can be computed using tools from topology, arithmetic groups give us a concrete way to calculate certain automorphic forms and thus L-functions associated to them. We emphasize that not all automorphic forms will appear in cohomology. For instance if $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$, then only the holomorphic forms of weight ≥ 2 are cohomological; the Maass forms and the weight 1 holomorphic forms are not. In any case, the cohomological automorphic forms are widely believed to have many relationships with Galois representations and motives.

To compute the cohomology $H^*(\Gamma; \mathscr{M})$, one typically instead computes the cohomology of the locally symmetric space attached to G and Γ. Let $K \subset G$ be a maximal compact subgroup, and let A_G be the split component. Then the globally symmetric space $D = G/KA_G$ is contractible, and if Γ is torsion free then the locally symmetric space $Y_\Gamma = \Gamma \backslash D$ is an Eilenberg–Mac Lane space for Γ. The module $\mathscr{M}$ defines a local coefficient system $\tilde{\mathscr{M}}$ on Y_Γ, and one can identify the group cohomology with $H^*(Y_\Gamma; \tilde{\mathscr{M}})$. Even if Γ has torsion, we can still identify group cohomology with the cohomology of the orbifold Y_Γ, since $\tilde{\mathscr{M}}$ comes from a complex representation of G. Thus we can use standard geometric techniques of algebraic topology to compute the cohomology.

To finish, we discuss a few important quantities attached to the cohomology. For simplicity we assume $\mathbf{G}$ is semisimple. First, although the cohomology a priori could occur in degrees $i = 0, \ldots, d = \dim Y_\Gamma$, the Borel–Serre vanishing theorem we have $H^i(Y_\Gamma; \tilde{\mathscr{M}}) = 0$ if $i > \nu := d - r_{\mathbb{Q}}(\mathbf{G})$, where $r_{\mathbb{Q}}(\mathbf{G})$ is the $\mathbb{Q}$-rank of $\mathbf{G}$. The number ν is called the *virtual cohomological dimension*. Second, the cusp cohomology will not appear in all cohomological degrees. Instead we have $H^i_{\mathrm{cusp}} = 0$ unless i lies in a symmetric interval about d of length $\delta := r_{\mathbb{C}}(G) - r_{\mathbb{C}}(K)$. The number δ is sometimes called the *deficiency* of G. If $G = \mathrm{SL}_3(\mathbb{R})$, for instance, then $d = 5$, $\nu = 3$, $\delta = 1$, and the cuspidal cohomology occurs in degrees $2, 3$.

6. Modular symbols for Siegel modular forms

Recently a number of large-scale computations have been carried out for Siegel modular forms:

- For spaces scalar-valued Siegel modular forms of genus two with full level, Raum [122] has computed bases of modular forms up to weight 172 in his investigation of Böcherer's Conjecture for Hecke eigenforms whose coefficients are not rational.
- For one- and two-dimensional spaces of vector-valued Siegel modular forms of genus two with full level, Faber and van der Geer [59] have computed systems of eigenvalues for $p \leq 37$ in order to verify a conjecture due to Harder [88] about congruences between Hecke eigenclasses of holomorphic modular forms vector-valued Siegel modular forms. These computations were recently extended in a way by Ghitza, Ryan and Sulon [76].
- Poor and Yuen [121] have computed a large number of Fourier coefficients of paramodular forms of weight two and prime level $p < 1000$. This was motivated by the Paramodular Conjecture [37]. Recently they have considered squarefree composite level.

Some smaller scale computations have been carried over the past decade or so, as well:

- Ibukiyama and his coauthors have computed the generating sets of rings of modular forms with small level [2, 91, 96]. These generators are expressed in terms of theta constants.
- Bergström–Faber–van der Geer have computed Hecke eigensystems for vector-valued Siegel modular forms of level 2 in order to verify Harder's conjecture in that setting [23].
- van Geemen and van Straten [73] have computed cusp forms of weight 3 on more exotic congruence subgroups of the Siegel modular group.

A feature of all the above computations is that the methods used to carry out the computations are *ad hoc*. For further investigation of Siegel modular forms, it is desirable to have a more systematic method that can be applied for a variety of weights and levels. For example, it seems difficult to explore generalizations of Harder's idea using existing methods.

We propose to investigate Siegel modular forms in genus 2 through the computation of the cohomology of Siegel modular threefolds. More precisely, let $\mathbf{G}$ be the semisimple group $\mathrm{Sp}_4/\mathbb{Q}$, $G = \mathbf{G}(\mathbb{R})$ the group of real points $\mathrm{Sp}_4(\mathbb{R})$, $K = U(2) \subset G$ a maximal compact subgroup, and $D = G/K$ the Siegel upper halfspace of genus 2. Let $\Gamma \subset \mathrm{Sp}_4(\mathbb{Z})$ be a congruence subgroup. The locally symmetric space $\Gamma\backslash D$ has complex dimension 3 and is known as a Siegel modular threefold. The cuspidal cohomology lives in degree $H^3(\Gamma\backslash D; \mathbb{C})$, and gives an incarnation of Siegel modular forms on subgroups of $\mathrm{Sp}_4(\mathbb{Z})$ of scalar weight 3. Higher weight forms can be found by using local coefficients $\mathscr{M}$. We remark that this "incarnation" of the forms allows one (in principle) to compute the eigenvalues of the Hecke eigenforms, but not their Fourier coefficients. Nevertheless, for investigating connections to arithmetic geometry, such as to abelian surfaces [37, 121] and hyperelliptic curves (Section 3.3), the Hecke eigenvalues are often exactly what one wants. One should also note that many Siegel modular forms of arithmetic interest are not cohomological, such as the weight 2 forms investigated in [121].

Our goal is to develop new topological techniques that supplement existing tools to compute $H^3(\Gamma\backslash D; \mathscr{M})$ and the Hecke action, and thus to compile substantial data for these forms. Since these techniques involve computing with combinatorial models for group cohomology in the spirit of [10–12, 81, 87], which build on the classical theory of modular symbols, this would constitute an analogue of modular symbols for Siegel modular forms. Work on this and allied activities would involve PIs Gunnells and Stein, as well as senior scientist Ash and LMFDB collaborator Ryan.

As in section 4.2, the computation divides into two parts. First one needs a cell decomposition of the modular threefold to compute the cohomology. Then one needs an algorithm to compute the Hecke action on the cohomology.

In general one does not know how to construct practical cell decompositions of locally symmetric spaces, except for the case of linear symmetric spaces. Such spaces have models in terms of quadratic forms and include GL_n over number fields and division algebras. For nonlinear spaces very few examples are known. Fortunately for us the case of Siegel modular threefolds has been successfully treated by McConnell–MacPherson [112, 113]. In particular we have an explicit cell decomposition, with cells indexed by nice combinatorial data (configurations of points in $\mathbf{P}^3$). Hence computing H^3 is easily within reach.

The Hecke operator computation is much more involved, and is where the new techniques will enter. The virtual cohomological dimension of Γ is $\nu = 4$, and we want to compute Hecke operators on H^3, one below this degree. Thus the situation is similar to that for $\Gamma \subset \mathrm{SL}_4(\mathbb{Z})$, and one would like to generalize the techniques of [79] to this setting. A necessary prerequisite is a modular symbol algorithm to compute the Hecke action on H^4, which has been accomplished in [80]. In particular [80] provides the first two groups in what should be the symplectic sharbly complex, a necessary ingredient for generalizing [79]. The first theoretical challenge is thus extending this to a full resolution of the symplectic Steinberg module. We have some clues about this resolution from the combinatorics of the McConnell–MacPherson cell decomposition; at the very least we want our resolution to contain $\mathrm{Sp}_4(\mathbb{Q})$-translates of these cells. Other sources of inspiration will be Tóth's work [134] on the Steinberg module for Chevalley groups, and Ash's recent new resolution of the Steinberg module for GL_n [7].

The Hecke eigenvalues obtained in this way can be used to produce the spin (degree 4) L-function of the Siegel modular form. This can be used to confirm modularity for hypergeometric L-functions (Section 9). Those connections will be incorporated into the LMFDB.

7. Artin representations

Let $G_{\mathbb{Q}}$ be the absolute Galois group of $\mathbb{Q}$. By an Artin representation, we mean a continuous homomorphism $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_n(\mathbb{C})$, where $\mathrm{GL}_n(\mathbb{C})$ has the discrete topology. Necessarily, ρ factors through $\mathrm{Gal}(K/\mathbb{Q})$ for some Galois number field K where the induced map $\mathrm{Gal}(K/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{C})$ is injective. We will denote this induced map as ρ, and refer to the field K as the Artin field of ρ. The associated Artin L-functions provide an interesting source of L-functions which can be used to factor Dedekind zeta-functions of number fields.

The LMFDB currently contains data for all irreducible Artin representations associated to the Galois closures of some of the number fields in the database. We propose several improvements.

First, we would like to strengthen the connection between Dedekind zeta-functions and Artin L-functions. Currently, the LMFDB shows the factorization of a Dedekind zeta-function as a product of Dirichlet L-functions for abelian extensions of $\mathbb{Q}$. We plan to extend this more broadly using Artin L-functions. This will require generating additional data on Artin representations to match the number field database, and making the association between non-Galois number fields and the associated irreducible Artin representations.

Second and more importantly, we would like to include complete tables of Artin representations in the LMFDB. If $\rho : \mathrm{Gal}(K/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{C})$ is an Artin representation and χ its character, let $f(\chi)$ denote its conductor. Given a finite group G, an irreducible character χ of G whose representation is faithful, and a bound B, let $\mathcal{A}(G, \chi, B)$ be the set of Artin representations ρ with Artin field K and character χ such that $f(\chi) \leq B$. This is a finite set.

In recent work with senior scientist Roberts, PI Jones has been developing ways of determining divisibility relations for discriminants of number fields [103]. Determining relations which are guaranteed to hold assuming all ramification is tame reduces to basic group theoretic computations, and we show that these often continue to hold in the presence of wild ramification. We plan to extend these ideas to conductors of Artin representations. Using this, one could then compute sets $\mathcal{A}(G, \chi, B)$ by computing corresponding sets of low degree number fields, with the transition between a bound on Artin conductors to a bound on field discriminants being optimal. Given the appropriate number fields, we can use the work of Dokchitser to generate data for the corresponding Artin representations and their L-functions.

Extending tables with bounded discriminant would allow us to compute $\mathcal{A}(G, \chi, B)$ in non-trivial examples, and with B large enough that the sets are non-empty. Preliminary estimates indicate that we will be able to carry this out for all $G \leq S_5$ and all faithful irreducible characters of these groups, as well as for some additional subgroups of larger symmetric groups.

Creating these tables will facilitate further research in a number of ways. For example, when studying questions related to Artin L-functions, one may want to first study the easiest examples, which often translates to restrictions on the conductor. What may be easiest for one line of study may be different from what is needed in another, so having complete lists increases the chance of having desirable examples. It will also allow the study of statistics for Artin representations.

One direction PI Jones and senior scientist Roberts will study is lower bounds for conductors for Artin representations. For many years, Odlyzko bounds [117] were the best available, which were then improved in [119]. These are optimal if one fixes only the degree of the representation, but more can be learned in the context above. If one fixes a group, a representation, and a signature, then we expect to be able to improve the lower bounds on the conductor of such a representation. Having tables which provide proven first examples then provide a good test for these bounds. If they are sharp, then there is hope that the new bounds will be good for larger groups.

These results will produce, in an initial range, reasonably complete information about Artin representations. This is an important step in our long-term goal of providing a comprehensive picture of Galois representations. The next steps are described in Section 11.

8. L-functions as tools for exploring other objects

We now describe projects which use analytic L-function techniques to answer questions with an arithmetic origin. This work will be led by PI Farmer and senior scientists Schmidt and Rubinstein.

8.1. **Finding bad factors.** There are many situations in which one has only partial information about an L-function, the missing information typically being the Euler factors at the bad primes and the level. Examples are Rankin-Selberg convolutions (Section 8.2), hyperelliptic curves (Section 3.1), and hypergeometric L-functions (Section 9). In the first two cases, in principle one can derive all the data about the L-function from the underlying object (for hyperelliptic curves this is under development). In the third case the L-function appears without first finding its underlying object, so there seems to be no direct way to determine the missing information.

It turns out that one can use the L-function itself to determine the missing information. That this should be possible is evident from the multiplicity one theorem: there is only one L-function with the given data at the good primes, therefore the missing data exists and is unique (provided that the data at the good primes actually corresponds to an L-function). The issue is to find an efficient way to determine the missing information. As described in Section 10.1, one can use the freedom in the approximate functional equation to write down a system of equations whose solution is the unknown coefficients. Since one can loop over the possible levels and signs of the functional equation, in theory this solves the problem.

At the AIM workshop *Hypergeometric motives* in June 2012, a working group led by PI Farmer and LMFDB collaborator Molin modified this approach to simultaneously determine the level and the unknown coefficients with one calculation. We illustrate in the case the unknown part of the level is a prime power p^K, where K, and the local factor $f_p(p^{-s})^{-1}$ at p, are to be determined. Writing $\Lambda(s) = \Lambda^*(s)p^{Ks/2}/f_p(p^{-s})$, where Λ^* is the "known" part of the completed L-function, we rearrange the functional equation to read:

$$\Lambda^*(s)p^{Ks/2}f_p(p^{s-1}) = \varepsilon\Lambda^*(1-s)p^{K(1-s)/2}f_p(p^s).$$

The key observation is that $p^{Ks/2}f_p(p^{s-1})$ is just another polynomial in $p^{s/2}$. So one can set up a system of linear equations, with the coefficient of that new polynomials as unknowns, and solve for the level and the local factor simultaneously – the only restriction is that one must guess an upper bound on the degree so that enough equations are generated. In its initial implementation the resulting system of equation was quite stable and good numerical results were obtained, for degree 4 L-functions of level up to 10^6.

We plan two outcomes in this direction. The first is a robust implementation of the method which can serve as a "black box" for determining the level and bad factors of a wide variety of L-functions. This can be used, for example, to generate data for L-functions of hyperelliptic curves – at least until we develop a complete analogue of Tate's algorithm for those objects. The second concerns operations on L-functions, which we discuss next.

8.2. **Operations on L-functions.** L-functions give rise to other L-functions via lifts and twists. These operations can be defined strictly in terms of the L-function. Symmetric powers, for example, have a simple expression in terms of the Satake parameters of the L-function (see equation (2.4) for a discussion of Satake parameters). But there is one important caveat: these operations are only easy to define in terms of the L-function at the good primes. At the bad primes (depending on how bad it is) you may need to look in detail at the underlying object in order to determine, for example, the level of the symmetric square.

But there is an alternative approach. By the multiplicity one theorem [67,105,124], an L-function is determined by its good Euler factors. Thus, the other Euler factors, and all the parameters in the functional equation, are determined (in some abstract sense) by the good factors. The point made in Section 8.1 above is that there is a practical method for determining these missing factors.

This opens the door to exploring the wide range of L-functions that can be built from other L-functions. For example, one could look at the Rankin-Selberg convolution of two L-functions and find its primitive factors, or check the functional equation and Riemann Hypothesis for symmetric and exterior powers; in essence, check the Langlands functoriality conjecture in vastly more cases than have been possible before.

We expect to have a robust implementation of these algorithms well before Workshop 5, when we address the issue of implementing L-function operations in the LMFDB.

9. Hypergeometric motives

Hypergeometric motives are certain one-parameter families of motives, which in one incarnation correspond to the classical hypergeometric differential equations with rational parameters. A prototypical example is the equation satisfied by the Gauss hypergeometric function $F(\frac{1}{2}, \frac{1}{2}, 1; t)$ whose associated L-function is that of the Legendre elliptic curve $y^2 = x(x-1)(x-t)$. Another example is the basic period of the Dwork pencil of quintic threefolds

$$x_1^5 + \cdots + x_5^5 - 5\psi x_1 \cdots x_5 = 0$$

that plays a prominent role in mirror symmetry. It gives rise to degree 4 L-functions.

Computing the L-function of these motives does not require the direct counting of points of varieties nor the calculation of a corresponding automorphic form. Instead it uses a p-adic formula of Katz for the trace of Frobenius, which is a finite version of a hypergeometric function.

Thus, this approach can easily produce an enormous number of interesting L-functions; a theorem of Katz gives precise recipes for the weight, the Γ-factors, and the Euler factors for the good primes. The computation of the Euler factors at the bad primes, the sign of the functional equation and the conductor/level are not yet fully understood and require further work. In many instances the motive is recognizable as related to familiar objects like, hyperelliptic curves, Siegel or Hilbert modular form, Artin L-functions, etc. Making such connections precise is an interesting and often computationally quite challenging goal.

The input is two r-tuples α, β of rational numbers, which correspond to the roots of cyclotomic polynomials, and a rational number t, from which one uses the hypergeometric function

$$F(\alpha, \beta; t) = \sum_{m \geq 0} \frac{\prod_{1 \leq j \leq r} (\alpha_j)_m}{\prod_{1 \leq j \leq r} (\beta_j)_m} t^m, \tag{9.1}$$

where $(\cdot)$ is the Pochammer symbol. The bad primes are those dividing the denominators of α_j or β_j, and the numerator or denominator of t.

Work in this area is advancing rapidly, led by Henri Cohen and senior scientist Fernando Rodriguez-Villegas. Of primary concern for this proposal is analytic techniques for completing the partial L-functions (Section 8.1). Unlike the other elements of this project, which involve primarily a research component followed by integration into LMFDB at a workshop, we see hypergeometric L-functions as a testing ground for how the LMFDB interacts with other endeavors. At Workshop 4 (in which Rodriguez-Villegas has agreed to be a co-organizer) we will incorporate hypergeometric motives into the LMFDB, work on automating the process of connecting those to the hyperelliptic curves, Siegel and Hilbert modular forms, and mod p Galois representations which are already in the LMFDB. Success in those efforts will be a major milestone in the LMFDB's path to becoming a comprehensive resource for the research community.

10. Conjectures and completeness of L-functions

In section 2 we described L-functions as a mechanism to help organize the LMFDB and to make visible the connections predicted by the Langlands program and modularity conjectures.

In this section we describe our efforts to systematically chart the landscape of L-functions and to methodically check the predictions of the Langlands program and other conjectures.

10.1. **Direct search for L-functions.** PI Farmer, Koutsoliotas, and Lemurell [62, 63] have described a practical method of generating an L-function given only its functional equation. That is, the L-function is created "out of nothing," without the need to first find an underlying object.The method takes the functional equation as input, and then creates a system of equations which the coefficients of the L-function must satisfy. It is not obvious that enough assumptions have been made to determine an L-function, but in practice it works in many cases. So far this approach has been successful in finding L-functions up to degree 4 and small level.

This method will be used to explore all possible functional equations up through degree 4, leading to several new types of L-functions which have never been seen before. It will then be a challenge to understand the source of the underlying objects which give rise to these L-functions.

10.2. **Identifying L-functions.** The PIs on this proposal are frequently contacted by mathematicians who have questions like: "Are these numbers the first few coefficients of a modular form?" Usually the answer is "I don't know." And the reason we don't know is that the available data is not organized well enough to answer such questions.

Akin to Neil Sloane's On-Line Encyclopedia of Integer Sequences, which provides a tool to search through known sequences of integers, we will develop tools that will allow researchers and students to see whether data they have computed is connected to objects in the LMFDB. Examples include checking if a number is a special value of an L-function, or a zero of an L-function, or if a sequence of numbers are the coefficients of an L-function or modular form.

We will develop tools needed to uncover these sorts of relationships. At the heart of this will be a master index of L-functions, that will contain, for billions of L-functions in our database, specific data that will allow one to identify it in a variety of ways.

As an initial step in this direction, we have, for millions of L-functions of degree 1 and hundreds of thousands of L-functions of degree 2, created a table which includes the first few non-trivial zeros of each of these L-functions, as well as the first few Dirichlet coefficients. We have also created a search tool that allows one to look for L-functions with, say, first zero in a specified interval, as well as tools for searching by Dirichlet coefficients.

However taking this to the next step of allowing one to uncover relationships between L-functions and their sources (i.e. arithmetic, algebraic, automorphic, etc.), requires more work. For one, we need to expand our table to include more zeros of the primitive L-functions that appear in the LMFDB. We will do so systematically, up to degree 4, and some reasonable bounds for their level and spectral parameters. These bounds will be dictated by the complexity of our algorithms, and our computing and storage resources.

Next, we will develop tools for making connections between L-functions. For example, to *experimentally* factor a given L-function into primitive L-functions, one could compare the non-trivial zeros of the given L-function (which can be computed by the LMFDB) to the zeros in our table. This would allow one to surmise its factorization into primitive L-functions, assuming its factors have been seen before.

As another example, for the problem of recognizing a floating point number as being related to the special value of an L-function, we plan to store, for each L-function in our table, a list of some of its special values as numerical approximations, as well as more precise information when available (ex, class numbers). To uncover relationships, we will combine our table with the PSLQ algorithm of Ferguson and Bailey to look for linear and algebraic relationships between the given number and our collection of special values.

11. Galois representations

Galois representations arise (at least conjecturally) from many objects in the LMFDB. The ℓ^k-torsion of an Abelian variety cuts out a Galois extension, which can be packaged together to give an ℓ-adic Galois representation. Many automorphic forms on $\mathrm{GL}(n)$ or $\mathrm{Sp}(2n)$ are expected to be associated to Galois representations. In fact, anything that has an algebraic L-function (in the terminology of Section 2.2) conjecturally comes from a Galois representation.

We will develop Galois representations as native objects in the LMFDB, allowing users to search and browse Galois representations. Each Galois representation will have its own home page with basic information and links to objects which are associated to that representation.

We propose to begin by considering Galois representations mod ℓ, that is, representations into $\mathrm{GL}_n(\mathbb{F}_\ell)$. This is timely because we will have many examples arising from ℓ-torsion on hyperelliptic curves (Section 3.1), which in this case are representations into $\mathrm{GSp}(\mathbb{F}_\ell)$ and from torsion classes in the cohomology of arithmetic groups (Section 12). Furthermore, we will already be in the midst of developing the connections to elliptic curves over number fields (Section 4), another plentiful source of such representations. This work will be led by PI Kedlaya, senior scientist Roberts, and LMFDB collaborator Voight.

A natural first step to organizing Galois representation is in terms of their image – for example $\mathrm{GL}_2(\mathbb{F}_5)$ or $\mathrm{Sp}_4(\mathbb{F}_2)$. This has connections to the Sato-Tate group and L-group discussed earlier in this proposal. We will provide a complete classification of the Galois representations with very small image. The first several cases have already been done, and a particular case we wish to study in detail is $\mathrm{GSp}_4(\mathbb{F}_3)$. This case is nontrivial, but should be within the range of what is possible.

In some cases mod ℓ representations are "torsion," which is to say that they yield a finite Galois group but do not lift to an ℓ-adic representation, and in some cases these are the only way we know these number fields. There are examples from weight 1 classical modular forms, GL_2 over imaginary quadratic fields and complex cubics, GL_3 and GL_4 classes, etc. The information available has the trace of Frobenius and a bit more data about the Galois representation (e.g. conductor); identifying the relationships between these classes–especially lifts–would be very interesting.

In the long run we expect the LMFDB to contain every class of Galois representation and to use Galois representations as an organizing principle – much like our current plan for L-functions. The preliminary efforts on Galois representations will lay the groundwork for that work in the future.

12. Torsion in the cohomology of arithmetic groups and arithmetic

As seen in previous sections, the complex cohomology of arithmetic groups is closely related to automorphic forms and (conjecturally) to arithmetic. It has only recently been realized that the *integral* cohomology of arithmetic groups has further importance, and in particular that the *torsion* in the cohomology should be connected to Galois representations; see for instance [5,6,8,13,17,93]. We emphasize that there are Galois representations that are apparently attached to torsion classes but not (to our knowledge) to complex classes (i.e., to classical automorphic forms). Thus such conjectures, if valid, will play a role in the Langlands philosophy. For our project, torsion classes represent new "automorphic-like" data of interest to number theorists and thus deserve inclusion in the LMFDB. Work on the projects in this section would be carried out by PIs Gunnells and Stein (for the cohomology/Hecke computations) and Jones (for identification of Galois representations). Also participating will be senior scientists Ash and Cremona.

It has also been recently realized that the torsion in the cohomology behaves very differently for various arithmetic groups. Certain groups (such as $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ or $\mathrm{Sp}_4(\mathbb{Z})$) are expected to have very little torsion in their cohomology, whereas others (such as $\Gamma \subset \mathrm{SL}_3(\mathbb{Z})$ or SL_2 over imaginary quadratic fields) are expected to have explosive growth in the torsion as a function of index. This torsion does not come from the finite subgroups of Γ and instead is predicted by *arithmetic*: heuristics of Bharghava [25] on the asymptotic number of number fields of fixed degree and bounded discriminant lead one to conclude that, if torsion in cohomology is to be related to mod p Galois representations, then there are simply "too many" mod p Galois representations in certain cases to be accounted for by the reduction mod p of characteristic 0 classes. For instance, let $\Gamma(N)$ be the principal congruence subgroup of $\mathrm{SL}_n(\mathbb{Z})$. Then the heuristics predict that the fraction of squarefree levels N for which there is an eigenclass in $H^*(\Gamma(N); \mathbb{F}_p)$ should be $\gg p^{-\delta}$, where δ is the deficiency (Section 5); see [22, §6] for more details. Hence when $\delta = 1$, one expects lots of torsion in the cohomology, with unbounded order as N increases.

Based on this, Bergeron–Venkatesh formulated the following conjecture [22]. For simplicity we assume our underlying algebraic group $\mathbf{G}$ is semisimple:

Let $\{\Gamma_n\}$ be a congruence tower of cocompact subgroups of Γ such that $\cap\Gamma_n = \{1\}$. Then

$$\lim_n \log |H_i(\Gamma; \mathscr{M})_{tors}| / [\Gamma : \Gamma_n]$$

exists for each i, and is nonzero unless $\delta = 1$ and $i = (\dim(G/K) - 1)/2$.

This conjecture (for noncocompact and in cohomology rather than homology) has been tested in recent work of PI Gunnells with Ash, Elbaz-Vincent, McConnell, Pollack, and Yasaki [9]. In particular, apparent exponential growth has been observed for congruence groups in $\mathrm{SL}_3(\mathbb{Z})$, $\mathrm{SL}_4(\mathbb{Z})$, and $\mathrm{GL}_2(\mathscr{O})$, where $\mathscr{O}$ is the integers of the cubic field of discriminant -23; for these examples $\delta = 1$. For other groups, such as $\mathrm{SL}_5(\mathbb{Z})$ or GL_2 over the field of fifth roots of unity), one still sees plenty of torsion in the cohomology, but not exponential growth. There are currently no predictions for the asymptotic behavior of the growth of the torsion when $\delta > 1$.

We have the following goals for the present proposal:

(1) Extend the computations in [9] to other groups and to as large levels as possible, and to nontrivial coefficient systems. In cases where the torsion is not expected to grow exponentially with the index, gather data to help formulate conjectures about the growth.
(2) Compute Hecke operators on the torsion classes and, as far as possible, determine the associated Galois representations.
(3) Collect all the generated data into the LMFDB. Of special interest here are the mod p Galois representations that do not appear to be attached to characteristic 0 classes.

13. Automorphic forms and arithmetic groups over function fields

Fix a prime p, let $q = p^f$, and let $\mathbb{F}$ be the finite field of q elements. Let $F/\mathbb{F}$ be a global function field, i.e., an extension of transcendence degree 1. Let S be a finite nonempty set of places of F of cardinality d and let $\mathscr{O} = \mathscr{O}_S \subset F$ be the corresponding ring of S-integers. This portion of the proposal concerns the study of automorphic forms on discrete groups over F via cohomology of arithmetic groups and analogues of modular symbols. Work will be carried out by PIs Gunnells, Stein, and Jones, with the support of senior scientists Ash, Böckle, and Bucur.

More precisely, let $G_0/\mathbb{F}$ be a simply-connected simple Chevalley group, let r be its rank, Γ a congruence subgroup of the $\mathscr{O}$-valued points of G_0, and G the Cartesian product over $v \in S$ of the F_v-valued points of G_0. Then a theorem of Harder [89] asserts that $H^i(\Gamma; \mathbb{C})$ vanishes if $i \notin \{0, rd\}$ and is finite-dimensional otherwise. The dimension of $H^{rd}(\Gamma; \mathbb{C})$ equals the multiplicity in the discrete spectrum of $L^2(\Gamma\backslash G)$ of the so-called special representation of G.

We can interpret the cohomology geometrically as follows. Each place $v \in S$ determines a simplicial complex X_v, the Bruhat–Tits building attached to G_0/F_v. The product $X = \prod_{v\in S} X_v$ is thus a polysimplicial complex on which Γ acts properly discontinuously. Hence we can identify $H^*(\Gamma; \mathbb{C}) \simeq H^*(\Gamma\backslash X; \mathbb{C})$.

Thus the situation is very analogous to the setup in Sections 4.2, 6, and 12, with the added benefit that the analogue of our symmetric spaces come equipped with a cellular structure. This suggests that the cohomology of Γ and thus certain automorphic forms on G can be studied using modular symbols and similar constructions. Indeed, Teitelbaum [133] has already studied GL_2 modular symbols when $F = \mathbb{F}[T]$ is the rational function field with $\mathscr{O} = \mathbb{F}[T]$. He published only a few examples of Hecke eigenforms, reproducing earlier work of Gekeler [74]. Recently Armana [3] has investigated the GL_2 symbols and has produced data for congruence subgroups of $\mathrm{GL}_2(\mathscr{O})$.

We propose to address the following goals:

(1) Build on Armana's work to produce a large table of GL_2 forms, and the corresponding elliptic curves for the appropriate forms. Relevant here will be work of Gekeler–Nonnengardt [75] and Carbone–Cobbs–Murray [39] on computing fundamental domains of Hecke congruence subgroups of $\mathrm{PGL}_2(\mathbb{F}[T])$.
(2) Extend GL_2 computations to finite extensions of $\mathbb{F}[T]$. The simplest cases beyond $\mathbb{F}[T]$ are the four rings that are coordinate rings of affine curves of a smooth projective curve minus one point that have strict class number one [58, 92].
(3) Extend Teitelbaum's theory to higher rank GL_n, and compute tables of eigenforms for $n = 3$ and possibly $n = 4$.
(4) Extend all these computations to nontrivial coefficients.
(5) Investigate the arithmetic significance of the torsion in the cohomology.

14. The need for a collaborative effort

Throughout this proposal we have described the many facets of L-functions, automorphic forms, and related objects, how they interact via the Langlands program and other major conjectures, and the many different arithmetic, analytic, and algebraic aspects that we will investigate from several different points of view. The goals of this proposal are by their very nature collaborative. They involve several rich subjects, with L-functions and Galois representations providing the major themes that tie the subjects together. We plan to study many kinds of L-functions including those associated to Artin representations, elliptic curves over number fields, hyperelliptic curves, Siegel and Hilbert modular forms, Maass forms for $GL(3)$ and $GL(4)$, and hypergeometric motives. We will also tabulate Galois representations, and automorphic forms over function fields. Our work will require expertise on analytic, arithmetic, algebraic, and computational aspects of these L-functions and their sources.

This project cannot be carried out without a concerted group effort. No single individual or collection of isolated individuals is capable of carrying out such a broad range of work. Furthermore, the LMFDB, which will emphasize how these different points of view and objects interrelate, requires the close collaboration between researchers from different backgrounds.

15. Timeline

We gather the many aspects of this project into a single timeline. There are two main components: theoretical work, which occurs throughout the year and primarily involves individuals and small groups, and LMFDB implimentation, which invlves a large group together at a workshop.

For each research activity we refer to the section of this proposal that provides more details about the mathematics and the roles of the PIs, Senior Scientists, and LMFDB collaborators.

For each workshop, for which we plan two each year, we indicate the **PIs**, *Senior scientists*, and current LMFDB collaborators who will play a leadership role. Each workshop is structured around "New" topics that do not currently appear in the LMFDB, and "Revision" in which significant enhancements are made to existing LMFDB functionality.

15.1. **Year 1 (summer 13 - summer 14).** Research topics: Elliptic curves over number fields (Sections 4.1 and 4.2). Classification of L-functions of degree ≤ 4 (Section 2.1). Modular symbols for Siegel modular forms (Section 6). Hyperelliptic curves of genus 2, bad factors (Section 3.1)). Hyperelliptic curves of genus 2, show list is complete (Section 3.3). Artin L-functions: complete list (Section 7). Numerically finding bad factors of L-functions (Section 8.1).

Workshop 1: (**Kedlaya**, **Farmer**, *Bucur, Sutherland*, *Schmidt*, Conrey, Ryan)

New: hyperelliptic curves

Revision: L-function structure and navigation, arithmetically normalized L-functions.

Workshop 2: (**Gunnells**, **Stein**, **Jones**, *Cremona, Mazur*, Dembele, Voight)

New: Elliptic curves over quadratic fields.

Revision: Hilbert modular forms.

15.2. **Year 2 (summer 14 - summer 15).** Research topics: Modular symbols for Siegel modular forms, continued (Section 6), Hyperelliptic curves of genus ≥ 3 (Section 9), Automate direct search for L-functions (Section 10.1), Galois representations mod p (Section 11).

Workshop 3: (**Gunnells**, **Jones**, *Roberts*, Dokchitser)

New: Galois representations

Revision 1: Artin representations and Artin L-functions Revision 2: Catalog of L-functions

Workshop 4: (**Kedlaya**, **Farmer**, **Gunnells**, *Rodriguez-Villegas*, *Roberts*)

New: Hypergeometric motives, and tools for completing L-functions.

Revision: Siegel modular forms

15.3. **Year 3 (summer 15 - summer 16).** Research topics: Torsion in the cohomology of arithmetic groups and connections to arithmetic. (Section 12), Function fields (Section 13), Operations on L-functions (Section 8.2), Hyperelliptic curves of genus ≥ 3, continued (Section 3.1)).

Workshop 5: (**Farmer**, *Schmidt, Rubinstein*)

New: Operations on L-functions

Revision: completing partial L-functions.

Workshop 6: (**Gunnells**, **Jones**, **Farmer**)

New: Function fields.

Revision: documentation and functionality – prepare LMFDB to serve as a long-term resource.

16. Prior Results

PI Farmer's past research includes results in random matrix theory and its connection to L-functions [48, 49, 60], zeros [47, 57, 61, 68], the relationship between L-functions and modular forms [46, 62, 64], and computation of L-functions and Maass forms [66, 69].

In the past five years PI Gunnells was supported by NSF grants DMS–0401525, DMS–0801214, and DMS–1101640. During that time he published papers on the cohomology of arithmetic groups [11–14, 81, 82, 86] Weyl group multiple Dirichlet series [33, 34, 41–45], Kazhdan–Lusztig cells [21, 78, 85], and higher-dimensional Dedekind sums [84].

PI Jones is currently funded under NSF grant DUE-1226081. Recent research has resulted in publications on number fields with prescribed ramification [55, 56, 99–104]. As PI on DUE-0340688 he helped organize a collection of 25,000 homework problems for the free web-based homework system WeBWorK, in use by hundreds of institutions.

PI Kedlaya's past research includes results on Sato-Tate groups of abelian surfaces [71, 107, 108], computing zeta functions using p-adic cohomology [1, 106], and numerical p-adic integration and applications [19].

PI Stein has received a SCREMS grant (DMS-0821725), an FRG (DMS-0757627), and two COMPMATH (DMS-0713225 and DMS-1015114) grants. He is currently a co-PI on a CCLI type 2 grant (DUE-1020378). Stein was mainly supported by ANTC (DMS-0653968) for the period 2007–2010, and this is the award most closely related to this proposal, so we report in more detail about this award. It resulted in numerous published papers on the arithmetic of elliptic curves, modular forms and abelian varieties, including [20, 31, 97, 98, 111, 118, 127], and one undergraduate number theory book [125]. It also resulted in the submitted papers [128, 130, 131] and the graduate level textbook [126] on the Birch and Swinnerton-Dyer conjecture.

17. Modes of collaboration and training

The mathematicians working on this project range from those who have been with the LMFDB from its inception, to others who are new to the project. Much of the proposed work is standard research, for which there is no particular need to mentor new individuals. However, the process of incorporating the fruits of this labor into the LMFDB website is somewhat different.

Changes to the website will take place during workshops. Here we organize participants into small groups to work on particular subareas. Any participant who is new to the project will be grouped with at least one LMFDB veteran. At first the veteran will do all of the interaction with the site, and as the workshop progresses, the new participant will progressively make more direct changes to the site.

We see the workshops as an opportunity to involve more people in this project. We actively encourage women, under-represented minorities, and researchers from primarily undergraduate institutions to participate in our workshops, and we will provide funding to support their participation.

The LMFDB, as a visible, active, online resource, provides an environment which supports a large virtual collaboration. We will investigate ways to implement a "news" or "what's new" feature on the LMFDB home page, so that people can easily find out about new materal on the site. As a followup to our current project, it may be worthwhile to explore how a social networking community could emerge through the LMFDB contributors.

18. Management Plan

Because of the large number of people on this project and the large-scale nature of the collaborative effort, we recognize two levels of organization: *logistical management* and *research management*. The PIs Farmer, Gunnells, Jones, Kedlaya, and Stein will manage both aspects of the project. Each brings skills that are needed for management. Farmer, Director of Programs at the American Institute of Mathematics, has 10 years of experience working with groups of people having various skills and interests, to forge a long-term plan that has significant impact on the research of the participants as well as on the direction of the field. All of the PIs have research interests that overlap several topics in the LMFDB and collectively cover the entire area. Farmer is chair of the Management Board of the LMFDB, Jones is Editor-in-Chief, and Kedlaya and Stein are members of the Management Board. Stein is the director of the Sage open source software development project, which has over 50 programmers from across the globe who have now written several hundred thousand lines of code for the project. Jones manages the code and data for the most complete database of number fields – data which is in the process of being moved to the LMFDB.

Jones and Stein will oversee the data and code in the LMFDB, developing procedures for the review of new contributions. Farmer will oversee the presentation and user experience aspects of the LMFDB website, ensuring that the site is a valuable resource for both experts and non-experts.

18.1. **Managing research.**

The PIs Farmer, Gunnells, Jones, Kedlaya, and Stein will oversee the general direction of this project. Each workshop will include a meeting in which the overall plan for the project is reviewed, progress to date is discussed, and the success of the workshops is evaluated. We have built in flexibility to the project that will allow us to provide more resources to aspects that are moving more slowly than expected, and to take advantage of new developments as they occur.

This project falls naturally into several sub-projects that make progress separately and in combination. As described in the main body of the proposal, each sub-project is overseen by a PI, and in most cases also a senior scientist. The PI is responsible for reporting on the activity of the sub-project and to quickly bring attention to any aspect that requires a larger group discussion.

The timeline of workshops serves as an objective measure of the progress on this project and will help to ensure that we stay on schedule.