Elaborator Reflection: Extending Idris in Idris

David Christiansen  
Indiana University Bloomington, USA  
davidchr@indiana.edu

Edwin Brady  
School of Computer Science, University of St Andrews, Scotland  
ceb10@st-andrews.ac.uk

Abstract
Many programming languages and proof assistants are defined by elaboration from a high-level language with a great deal of implicit information to a highly explicit core language. In many advanced languages, these elaboration facilities contain powerful tools for program construction, but these tools are rarely designed to be repurposed by users. We describe elaborator reflection, a paradigm for metaprogramming in which the elaboration machinery is made directly available to metaprograms, as well as a concrete realization of elaborator reflection in Idris, a functional language with full dependent types. We demonstrate the applicability of Idris’s reflected elaboration framework to a number of realistic problems, we discuss the motivation for the specific features of its design, and we explore the broader meaning of elaborator reflection as it can relate to other languages.

Categories and Subject Descriptors  D.3.3 [Programming Languages]: Language Constructs and Features

Keywords Metaprogramming, dependent types, elaboration

1. Introduction
Writing programs that write programs, or metaprogramming, allows ordinary programmers to seize privileges that were once reserved for language implementers. This is especially true when metaprogrammers are able to extend the languages that they use to express themselves. Language extensions can eliminate the need for tedious boilerplate code, they can enable domain-specific idioms, notation and design patterns to be expressed directly, and they can enable optimizations that were not foreseen by the compiler’s authors. In languages based on dependent type theory, metaprograms are widely used for automating the construction of proofs as well as programs.

Idris is a pure functional language with full dependent types. Idris is type checked in two stages: first, a metaprogram known as an elaborator translates the user-accessible language into a much smaller core language, known as TT, after which the TT program is type checked independently with a much simpler type checker.

Elaboration allows Idris to have high-level conveniences whose safety is guaranteed by the core language’s type checker. These conveniences include implicit arguments that are found by the compiler, type classes,1 type-driven disambiguation of names, and local definitions. The core language has none of these. Elaborating each of these features affects the elaboration of the others. For example, resolving a type class might cause the value of an implicit argument to be discovered, which itself provides type information that enables a name to be disambiguated, which then might cause still more implicit arguments to be solved. This means that the elaboration procedures for each of these features must be able to share information and that they cannot be written in isolation.

A difficult metaprogramming problem calls for a powerful solution. Taking a cue from successful metaprogramming systems for proof automation, Brady (2013) based the Idris elaborator on the design of tactic-based interactive proof assistants, embedding proof tactics in a Haskell monad. The elaborator is then a function from high-level Idris abstract syntax trees to programs in this tactic language, which construct the desired TT program when run. The abstractions and effects in the tactic language, including holes to be filled in and unification problems yet to be resolved, enable the elaborators for each language feature to communicate without needing to know about each others’ implementation details. Additionally, the tactic-based elaborator has proven to be robust in the face of extensions that were not conceived of at the time of its initial implementation, such as quasiquotation (Christiansen 2014). Unfortunately, the elaborator monad is part of the underlying language implementation, and modifications and extensions require changing the compiler, erecting a solid barrier between language users and language implementers.

In this paper, we introduce elaborator reflection, where Idris’s elaboration framework is realized as a primitive monad in Idris itself. This empowers ordinary Idris users to write elaboration scripts in Idris, and running these elaboration scripts is as easy as type checking an Idris file, rather than rebuilding the compiler. With this new power at our fingertips, we discovered that Idris’s elaborator monad turns out to be useful for more than just elaborating high level Idris programs. We extended Idris with the ability to write separate termination proofs for general recursive functions by automating the Bove-Capretta transformation (Bove and Capretta 2005), we automated the proofs of a number of the lemmas in Idris’s standard library, and we demonstrated how to generate data types that match the schemas of external data sources based on Idris’s type providers (Christiansen 2013), all without leaving the cozy confines of our favorite language.

Contributions
This paper makes the following contributions:

• We present elaborator reflection, a new paradigm for metaprogramming in which metaprograms output a simpler core lan-

---

1 Type classes are called “interfaces” in recent versions of Idris, but in the present paper we will use the Haskell terminology.
guage and have access to the elaboration facilities of the high-
level language’s implementation.

- We demonstrate a concrete realization of elaborator reflection in
  the dependently typed functional language Idris.
- We demonstrate that Idris’s elaborator reflection is general
  enough to support diverse metaprogramming tasks, including
type class derivation, proof automation, interactive code genera-
tion, and even reimplementation of compiler features as libraries.

Additionally, elaborator reflection may provide an alternative route
to self-hosting. We comment on this possibility in Section 7.

2. Motivating Examples

Elaborator reflection allows us to write metaprograms, describing
mechanical procedures for automatically generating programs in
Idris’s core language. In this section we briefly describe two exam-
ple systems which will benefit from elaborator reflection: automatic
generation of termination predicates to help reason about totality; and
automation of a family of proofs. Here, we describe the problems;
later, in Section 5, we will describe how to solve these problems and
others using elaborator reflection.

Before discussing more realistic examples, it can be instructive
to see a small example of elaborator reflection in action. Firstly,
we construct a metaprogram that will implement a monomorphic
identity function at a type determined by the context in which it is
invoked:

```elab
mkId : Elab ()
mkId = do x <- gensym "x"
  attack
  intro x
  fill (Var x)
  solve
  solve
```

The type annotation Elab () declares that mkId is a metaprogram
that uses elaborator reflection. The use of do notation desugars to the
monadic bind operator, just as in Haskell. The first step in generating
the function is to construct a unique name for its argument using
gensym, which returns a fresh name. The next tactic, attack, is
used to delimit fresh scopes. It is described in detail in Section 4.1.
The lambda term is constructed using intro, which automatically
populates the type annotation of the lambda from the expected
function type. The body of the lambda is filled with a reference to
its bound variable x. Because fill only places a provisional term
in the focused hole, the solve tactic is used to mark that provisional
term as a complete solution. Finally, solve is used again, this time
to mark the scope introduced by attack as completed.

The metaprogram is invoked using the syntax %runElab mkId,
which tells Idris to build a term by applying the elaborator mkId. It
can be used in many different contexts that expect identity functions:

```elab
idNat : Nat -> Nat
idNat = %runElab mkId
idUnit : () -> ()
idUnit = %runElab mkId
idString : String -> String
idString = %runElab mkId

theAnswer : Int
theAnswer = (%runElab mkId) 42
```

We will return to a detailed description of elaborator reflection
soon. First, we present two interesting metaprogramming problems.

2.1 Termination Predicates

Idris is primarily a programming language rather than a proof
assistant. As such, it admits programs that it cannot determine to be
total functions and programmers do not, in general, need to write
termination proofs. To maintain its consistency as a logic, non-total
programs are not reduced during type checking, but are otherwise
compiled normally.

However, a user who has written a total function that does not pass
the termination checker may later decide to use it in a type,
either to prove some property about it or to be able to exploit its
reduction behavior. To use this function in a type, the user must
rewrite it to contain an explicit termination argument. Unfortunately,
such a modification typically has a drastic negative impact on the
readability of the function, commingling the administrivia of the
termination argument with the interesting flow of the algorithm
being implemented. Ideally, a user would specify an algorithm, and
then receive a separate proof obligation for its termination condition,
separating the mundane bookkeeping of the transformation from the
interesting termination argument.

For example, a programmer might write a functional analogue of
quicksort:

```elab
quicksort : List Nat -> List Nat
quicksort [] = []
quicksort (x::xs) =
quicksort (filter (< x) xs) ++
x :: quicksort (filter (>= x) xs)
```

Then, the programmer might decide that they want a total version of
quicksort. Ideally, it should be sufficient to write:

```elab
make quicksort total by qsTotalityProof
```

where qsTotalityProof becomes the name of a proof obligation for a
termination argument.

Rather than hard-coding such an interactive feature by directly
modifying our language implementation, we can use the reflected
elaborator to implement it as a library. Bove and Capretta (2005)
demonstrated a means of rewriting a large class of functional
programs that use general recursion into a form acceptable to
proof assistants. The Bove-Capretta transformation of a program
consists of an inductive family that represents the call graph of the
program, indexed by the program’s arguments, and a function that is
structurally recursive over these call graphs. The inductive family is
typically referred to as an accessibility predicate. For quicksort,
the accessibility predicate is:

```elab
data QSAcc : List Nat -> Type where
  QSAccCase0 : QSAcc []
  QSAccCase1 : QSAcc (filter (< x) xs) ->
                QSAcc (filter (>= x) xs) ->
                QSAcc (x :: xs)
```

The well-foundedness of the accessibility predicate can serve as
evidence that the program will return a value in finite time for a
particular set of arguments that have been proven to be accessible,
and it becomes possible to prove that the function is total by showing
that all possible arguments are accessible. For quicksort, we write
a helper function qs’ by recursion over a QSAcc predicate:

```elab
qs' : (xs : List Nat) -> QSAcc xs -> List Nat
qs' [] QSAccCase0 = []
qs' (x :: xs) (QSAccCase1 l r) =
  qs' (filter (< x) xs) l ++
  x :: qs' (filter (>= x) xs) r
```

Additionally, accessibility predicates are collapsible (Brady et al.
2003) and can therefore be eliminated from the run-time behavior of
the program, because the particular constructor of the accessibility
Elaboration of a high level Idris program into TT involves traversing the syntax of the high level program, incrementally constructing a TT term. During elaboration, the TT term may contain holes, which stand for parts of the program which have not yet been elaborated. Elaboration is type-directed, meaning that the elaborator always knows the type of the term it is constructing. The elaborator may also generate unification constraints between the type of the
term it is constructing, and the required type. Such constraints may often need elaboration to make further progress before they can be resolved. There are, therefore, several pieces of state which need to be maintained during elaboration:

- a **goal type**, which is a $\text{TT}$ term representing the type to be inhabited by the current elaboration procedure;
- an incrementally constructed **proof term**, which is a $\text{TT}$ term that must inhabit the goal type;
- a **hole queue** containing the holes and guesses that remain in the proof term;
- a collection of open unification problems;
- a supply of fresh names; and miscellaneous pieces of state to implement other internal operations.

We refer to these collectively as the **proof state**.

Instead of maintaining a separate metavariable context that tracks the unique names, scopes, expected types, and equality constraints related to the holes in the term, the Idris elaborator equips $\text{TT}$ with two additional binding forms, using the already-existing scoping and equality mechanisms for terms and contexts to represent metavariables in-place. This style was pioneered by McBride’s (1999) Oleg proof assistant. We refer to this extended core language as the **development calculus** $\text{TT}_{\text{dev}}$. Figure 2 shows the extensions to the syntax of hole and guess binders in $\text{TT}_{\text{dev}}$.

In $\text{TT}_{\text{dev}}$, the term $\exists x : t. e$ is the term $e$ with an additional bound variable $x$. The typing rules for hole binders assign the type $t$ to $x$ in its scope, demanding no further justification because further elaboration will be required to determine the actual term of type $t$ to be inserted. Because we are working modulo Idris’s definitional equality which includes $\beta$-conversion, directly substituting a term for the hole $x$ might trigger a sequence of reductions that interferes with further processing that we intend to perform. Thus, instead of directly substituting a term for a hole variable, we first construct a temporary binder that resembles a let but is computationally inert. This placeholder binder is called a **guess**. If the term $e'$ is inserted into the hole $x$ above, the resulting guess is written $\exists x \approx e'. e$. Then, when concerns about the precise structure of the body of the guess binder are no longer relevant, the guess can be substituted for the variable that it binds.

During elaboration, the names of holes and guesses are unique, so that the hole queue can contain just the names. When the hole queue is non-empty, elaboration is said to be **focused** at the hole or guess whose name is at the head of the queue. Many tactics implicitly modify the focused hole. When a guess is substituted for the variable that it binds, focus proceeds to the next hole or guess in the queue.

The Idris elaborator interprets each high-level Idris construct as a computation in an elaborator monad:

```
data Elab a -- abstract
```

Elaboration scripts are written using a collection of **basic tactics**, which are small, reusable building blocks for proof term construction. Some of the most commonly used tactics are:

- **claim**, which wraps the focused subterm with a new hole binding for a given name and type;
- **fill**, which converts a hole into a guess given an appropriately-typed term; and
- **solve**, which substitutes the focused guess in its scope and focuses on the next term.

We will describe these and other tactics in detail in Section 4.1. The Idris elaborator (and hence the Elab data type) is written in Haskell, so any features of the language are written as programs in the Elab monad in Haskell. The purpose of elaborator reflection is to expose the Elab monad in Idris programs, thus allowing Idris extensions to be written directly in Idris itself. To use this, we need an Idris representation of the Elab representation of $\text{TT}$ programs, and an Idris API for the elaborator monad.

### 3.3 Reflection Data Types

Idris’s reflection API uses a number of data types for representing terms and definitions, as shown in Figure 3.

**Figure 2. $\text{TT}_{\text{dev}}$ extensions**

<table>
<thead>
<tr>
<th>$b ::=$</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\exists x : t$ (hole binding)</td>
</tr>
<tr>
<td></td>
<td>$\exists x \approx t : t$ (guess)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\text{TT}_{\text{dev}}$ extensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>term it is constructing, and the required type. Such constraints may often need elaboration to make further progress before they can be resolved. There are, therefore, several pieces of state which need to be maintained during elaboration:</td>
</tr>
<tr>
<td>- a <strong>goal type</strong>, which is a $\text{TT}$ term representing the type to be inhabited by the current elaboration procedure;</td>
</tr>
<tr>
<td>- an incrementally constructed <strong>proof term</strong>, which is a $\text{TT}$ term that must inhabit the goal type;</td>
</tr>
<tr>
<td>- a <strong>hole queue</strong> containing the holes and guesses that remain in the proof term;</td>
</tr>
<tr>
<td>- a collection of open unification problems;</td>
</tr>
<tr>
<td>- a supply of fresh names; and miscellaneous pieces of state to implement other internal operations.</td>
</tr>
</tbody>
</table>

We refer to these collectively as the **proof state**.

Instead of maintaining a separate metavariable context that tracks the unique names, scopes, expected types, and equality constraints related to the holes in the term, the Idris elaborator equips $\text{TT}$ with two additional binding forms, using the already-existing scoping and equality mechanisms for terms and contexts to represent metavariables in-place. This style was pioneered by McBride’s (1999) Oleg proof assistant. We refer to this extended core language as the **development calculus** $\text{TT}_{\text{dev}}$. Figure 2 shows the extensions to the syntax of hole and guess binders in $\text{TT}_{\text{dev}}$.

In $\text{TT}_{\text{dev}}$, the term $\exists x : t. e$ is the term $e$ with an additional bound variable $x$. The typing rules for hole binders assign the type $t$ to $x$ in its scope, demanding no further justification because further elaboration will be required to determine the actual term of type $t$ to be inserted. Because we are working modulo Idris’s definitional equality which includes $\beta$-conversion, directly substituting a term for the hole $x$ might trigger a sequence of reductions that interferes with further processing that we intend to perform. Thus, instead of directly substituting a term for a hole variable, we first construct a temporary binder that resembles a let but is computationally inert. This placeholder binder is called a **guess**. If the term $e'$ is inserted into the hole $x$ above, the resulting guess is written $\exists x \approx e'. e$. Then, when concerns about the precise structure of the body of the guess binder are no longer relevant, the guess can be substituted for the variable that it binds.

During elaboration, the names of holes and guesses are unique, so that the hole queue can contain just the names. When the hole queue is non-empty, elaboration is said to be **focused** at the hole or guess whose name is at the head of the queue. Many tactics implicitly modify the focused hole. When a guess is substituted for the variable that it binds, focus proceeds to the next hole or guess in the queue.

The Idris elaborator interprets each high-level Idris construct as a computation in an elaborator monad:

```
data Elab a -- abstract
```

Elaboration scripts are written using a collection of **basic tactics**, which are small, reusable building blocks for proof term construction. Some of the most commonly used tactics are:

- **claim**, which wraps the focused subterm with a new hole binding for a given name and type;
- **fill**, which converts a hole into a guess given an appropriately-typed term; and
- **solve**, which substitutes the focused guess in its scope and focuses on the next term.

We will describe these and other tactics in detail in Section 4.1. The Idris elaborator (and hence the Elab data type) is written in Haskell, so any features of the language are written as programs in the Elab monad in Haskell. The purpose of elaborator reflection is to expose the Elab monad in Idris programs, thus allowing Idris extensions to be written directly in Idris itself. To use this, we need an Idris representation of the Elab representation of $\text{TT}$ programs, and an Idris API for the elaborator monad.

### 3.3 Reflection Data Types

Idris’s reflection API uses a number of data types for representing terms and definitions, as shown in Figure 3.

```
-- Variable names
data TTName = ...
```

-- Constants
data Const = Int | Str String | ...

-- Binders
data Binder : Type -> Type where
  Lam : (ty : a) -> Binder a
  Pi : (ty, kind : a) -> Binder a
  Let : (ty, val : a) -> Binder a
  PVar : (ty : a) -> Binder a
  Hole : (ty : a) -> Binder a
  Guess : (ty, val : a) -> Binder a
```

```
-- Terms which have not yet been typechecked
data Raw = Var TTName
         | RBind TTName (Binder Raw) Raw
         | RApp Raw Raw
         | RConstant Const
         | RType
```

```
-- Well typed, de Bruijn indexed terms
data TT = P NameType TTName TT
         | V Int
         | Bind TTName (Binder TT) TT
         | App TT TT
         | TConst Const
         | TType TTUExp
```

**Figure 3. Outline of the reflected $\text{TT}$ data type**

The core language is represented by two separate data types: $\text{Raw}$ and $\text{TT}$. $\text{Raw}$ is used to represent terms that are to be submitted to the type checker, while $\text{TT}$ represents terms produced by the type checker. The differences between them are that $\text{Raw}$ does not mention universe levels and it does not distinguish between locally and globally bound names, with a single constructor for variable references. $\text{TT}$, on the other hand, decorates the type of types with information about universe levels and it uses a locally nameless representation, with de Bruijn indices for bound variables. Additionally, $\text{TT}$ contains local caches of the types of global variables, allowing $\text{TT}$ terms to be understood in isolation. Both representations of the core language have a uniform representation of binders, implemented by the data type $\text{Binder}$.

Idris names are represented using the $\text{TTName}$ data type. In addition to names that users might write, with or without namespace
qualifiers, \texttt{TTName} includes constructors for machine-generated unique names as well as names that result from inner scopes, such as where blocks. Additionally, it includes a constructor that is reserved for names generated by user elaboration scripts so that they can define globally-unique top-level names for helper definitions.

While the majority of the data types involved in elaborator reflection are used to represent details about the core Idris language, a few of them mention features of high-level Idris. In particular, top-level type signatures are represented by a list of \texttt{FunArg} followed by a return type, rather than by a dependent function type in the core theory:

\begin{verbatim}
data Plicity =
     Explicit | Implicit | Constraint

record FunArg where
     constructor MkFunArg
           name : TTName
           type : Raw
           plicity : Plicity
           erasure : Erasure

This allows reflection scripts to specify whether each argument is to be passed implicitly, explicitly, or using type class resolution, as well as whether the arguments are intended to be runtime-relevant. By looking up this specification in the global context, elaboration scripts can determine the intended calling conventions of Idris functions and attempt to infer implicit arguments as users would expect. Additionally, record types such as \texttt{TyDecl}, \texttt{DataDefn}, and \texttt{FunDefn} describe new type declarations, data declarations, and pattern-matching functions to be added to the global context.

Finally, reflected error messages are represented as lists of the type \texttt{ErrorReportPart}, which contains constructors for embedding strings, Idris terms, names, and even other error messages in an error. This allows error messages that result from metaprograms to be rendered using Idris’s pretty printer and to take advantage of some of Idris’s interactive features.

### 3.4 Quasiquote

Working directly with the \texttt{Raw} and \texttt{TT} data types when writing metaprograms in Idris has many of the same disadvantages as programming directly in TT. Christiansen (2014) described how to use Idris’s elaborator to translate quotations of high-level Idris terms into the corresponding reified TT terms. While it is certainly possible to use the reflected elaborator without quotations, they are a convenient interface and are used extensively in this paper. The expression syntax:

\`
\langle \texttt{e} \rangle
``

refers to a quotation of the expression \texttt{e} for which a type must be inferable, while the syntax:

\`
\langle \texttt{e : t} \rangle
``

represents a quotation of a term at some particular type (Idris does not have a native type annotation syntax for expressions, as it is not necessary elsewhere). Idris quotations are in fact quasiquotations, which means that sub-regions of a quotation can be evaluated rather than quoted. These sub-regions, called \textit{antiquotations}, are prefixed with a tilde (\`
). Quotations can be used both as ordinary expressions and as patterns, with antiquotations in patterns containing patterns. Additionally, both the TT and \texttt{Raw} data types can be quoted, and Idris uses the expected type to determine which representation to produce.

Idris also has a quotation syntax for \texttt{TTName}. Because Idris supports pervasive type-disambiguated overloading, there is a distinction between names as they are written in a source file and the full, qualified version that is eventually discovered during elaboration. The syntax \`\{n\}` is an unresolved quotation of the name \texttt{n} — that is, it is a quotation of the name \texttt{n} precisely as written. The similar syntax \`\{n\}` is a resolved quotation, where the name \texttt{n} is treated as a reference to a unique name in scope, and expanded with namespace qualifiers prior to reification as a \texttt{TTName}. If \texttt{n} cannot be uniquely resolved, then \`\{n\}` will result in an error. Typically, unresolved names should be used for new things being defined by a metaprogram, while resolved names should be used for existing definitions that the metaprogram will consult.

\section{Elaborator Reflection}

An Idris program consists of a series of declarations, which can be data types, records, functions, type classes, and instances, among others. Elaborator reflection adds a new declaration form:

\texttt{\%runElab e}

Here, \texttt{e} is an Idris expression. This will first elaborate \texttt{e} with the goal type \texttt{Elab \{}\texttt{\}. If successful, the result is then run as an \textit{elaboration script} by a built-in interpreter that assigns elaboration effects to a collection of primitive elaboration scripts. This can be understood by analogy to run-time side effects. Like Haskell, Idris has a family \texttt{IO} of computations that are assigned effectful semantics by the language runtime: a program with type \texttt{IO a} produces a value of type \texttt{a}, potentially having used side effects to do so. Analogously, a program in \texttt{Elab a} produces a value of type \texttt{a}, potentially having used side effects in the elaborator. While \texttt{IO} actions are run implicitly by being assigned the name \texttt{main}, actions in \texttt{Elab} are run explicitly using the \texttt{\%runElab} syntax.

Similarly, elaborator reflection adds a new expression form:

\texttt{\%runElab e}

This works in exactly the same way as the declaration form, except that the elaboration script is run in the context of expression elaboration.

\subsection{Elaborator Operations}

\texttt{Elab} is an instance of the \texttt{Functor}, \texttt{Applicative}, and \texttt{Monad} type classes. Additionally, a backtracking failure-handling operator is used to implement \texttt{Alternative}, allowing the standard Idris choice operator \texttt{<>} to be used to express error recovery. Errors can occur when tactics are misused, or they can be explicitly signaled by the \texttt{fail} tactic, which takes a reflected error message as an argument.

The additional effects available in \texttt{Elab} largely correspond to the low-level tactics mentioned in Section 3.2, described in detail by Brady (2013). While Brady’s article distinguishes between expression and definition elaborators, with the latter invoking the former, the reflected elaborator allows the mixing of both forms of operation and thus expands these tactics with operators for both querying and extending the global context. This is to allow the same elaboration code to be used in both expression and definition contexts, as well as to allow elaboration scripts to produce expressions that depend on auxiliary definitions or data types. In this section, we describe the most important primitives in detail. A comprehensive description is available in the first author’s Ph.D. thesis (Christiansen 2016).

As described in Section 3, the elaboration monad has a state that includes a proof term with holes in it, a queue containing these holes, and a collection of open unification constraints. The following primitive operators in \texttt{Elab} manipulate or query the proof state:

\texttt{gensym : (hint : \text{String}) \to \text{Elab TTName}}

Produce a unique name, which is useful when establishing a new binder.
focus : (h : TTName) -> Elab ()

Move the focus to the hole h, bringing it to the front of the hole queue, or fail if the hole h does not exist.

unfocus : (h : TTName) -> Elab ()

Move the hole h to the end of the hole queue, or fail if the h does not exist.

clear : (e : Raw) -> Elab ()

The following queries against the global environment are provided as primitives in Elab:

getEnv : Elab (List (TTName, Binder TT))

Look up the lexical scope at the focused hole, or fail if there are no holes.

lookupTy : (n : TTName) -> Elab (List (TTName, NameType, TT))

Look up the core-language types of every overloading of a name n in the global context. The NameType indicates whether the name is a user-defined constant, a data constructor, or a type constructor.

lookupDatatype : (n : TTName) -> Elab (List Datatype)

Look up the reified definitions of all data types whose names are overloading of n.

lookupFunDefn : (n : TTName) -> Elab (List (FunDefn TT))

Look up the reified definitions of all top-level bindings whose names are overloading of the n.

...
Section 2.1. The operators for extending the global context are:

- `declareDatatype :: TyDecl -> Elab ()`: Declare a type constructor for an inductive family to the global context.
- `defineDatatype :: DataDefn -> Elab ()`: Define an inductive family by its list of constructors.
- `lookupArgs :: (n : TTName) -> Elab (List (TTName, List FunArg, Raw))`: Look up the high-level Idris calling conventions (that is, which arguments are to be found implicitly) for all overloads of n.
- `norm : (env : List (TTName, Binder TT)) -> (tm : TT) -> Elab TT`: Invoke the evaluator to normalize tm relative to the global context and the lexical context env.
- `check : (env : List (TTName, Binder TT)) -> (tm : Raw) -> Elab (TT, TT)`: Invoke the type checker to produce a fully-annotated term and its type, considered with respect to a particular lexical environment env.
- `appendDecl = TyDecl -> Elab ()`: Add a declaration to the global context.
- `declareType :: TyDecl -> Elab ()`: Add a type declaration.
- `defineFunction :: FunDefn Raw -> Elab ()`: Use a sequence of pattern-matching clauses to define a function. As a special case, zero-argument definitions produce top-level bindings of constants.
- `metavar : TTName -> Elab ()`: Solve the current hole as if it were a top-level Idris named hole. Not to be confused with the holes in `TT_decl`, which are an implementation technique for elaborators, Idris’s holes represent unfinished user programs. These holes can later be solved using all of Idris’s standard interactive features. This tactic allows metaprograms to delegate to human intelligence when necessary.

As an example of tactic-driven definitions, the Idris data type definition:

```
class instance dictionaries, Idris also contains a general-purpose proof search that attempts to use the constructors of inductive types to solve a goal. Proof search is invoked during elaboration when solving for implicit arguments decorated with the `?` modifier as
Tactic scripts are notorious for being difficult to read without tool support, and tactic scripts built with Idris’s elaborator reflection are no exception. Thus, Idris provides an interactive shell in which Elab actions can be executed one at a time, replacing a previous special-purpose tactic language. This interactive shell provides programmers with immediate feedback as to the effect of their actions. Additionally, it allows them to inspect values that result from Elab computations, such as the internal representations of reified definitions. At the end of an interactive elaboration session, Idris will optionally append the elaboration script to the current source file, so that loading it will execute the tactic script.

Additional tools can be built on top of the elaboration shell. For example, the Emacs mode for Idris provides a proof script development environment, reminiscent of Proof General (Aspinall 2000), in which programmers can interactively step forwards and backwards a tactic script, inspecting the proof state at each step.

4.3 Staging Issues

From the perspective of an Idris user who is not using elaborator reflection, the elaboration of an Idris program proceeds from top to bottom. Names must be declared before use. Idris additionally supports mutually recursive definitions. Some expressions, such as case expressions, require additional top-level definitions to be produced after elaboration, but these helper definitions are not accessible to other parts of the program, which means that users can be blissfully unaware of this elaboration order.

This account of name and definition visibility becomes more challenging in the presence of elaborator reflection. Like Template Haskell, Idris’s elaborator reflection introduces tricky staging issues. Metaprograms in the elaborator can have side effects such as defining new names, filling holes, or even aborting elaboration.

For example, lists are elaborated from left to right. This can be observed using the following declaration:

```
x : List Nat
xs = let x = 8 Z
     in [xrunElab (do declareType valDecl
                      fill `(Z)
                      solve),
           x + val]
```

The name val in the second position in the list is not bound until the elaborator script in the first position has executed. If list elaboration were to proceed from right to left, then val would be unbound, so it is also possible to observe that the value of valDecl has been elaborated prior to the execution of the elaborator script, because it is available. Furthermore, simply lifting Elab expressions to the top level and elaborating them prior to the elaboration of the context in which they are executed is not sufficient: if it were computed via a case expression, the script would fail, because the computational behavior of the case expression is not elaborated until after xs has been elaborated. In the future, we hope to make the dependencies between elaboration stages more explicit.

5. Using the Reflected Elaborator

A metaprogramming system is of no use if it cannot be used to write interesting metaprograms. In this section, we deliver on our promises from Section 2 and describe additional experiences with the reflected elaborator.

5.1 Termination Predicates

In Section 2.1 we showed the following general recursive definition of quicksort:

```
quicksort : List Nat -> List Nat
quicksort [] = []
quicksort (x::xs) =
  quicksort (filter (< x) xs) ++
  x :: quicksort (filter (>= x) xs)
```

Idris cannot automatically determine that this function terminates. We can, however, show that it terminates using an accessibility predicate. Using elaborator reflection, we can automatically generate accessibility predicates by directly adapting Bove and Capretta’s method (2005) to rewrite Idris programs that make use of general recursion into a pair of an accessibility predicate and a function defined by recursion over the predicate. We begin by declaring the type of the elaboration procedure:

```
bc : (inFun, outFun, accPred : TTNam) -> Elab ()
bC inFun outFun accPred = do
```

The procedure bc takes three arguments: inFun, which is the partial program to be rewritten, outFun, which is the name to use for the rewritten function, and accPred, which is the name to use for the accessibility predicate. In the do block, the first step is to check that the provided name uniquely identifies a function:
We declare the rewritten function's type and the type constructor.

The new argument

- Exact

%runElab (bc `{quicksort} `{{qs'}} `{{QSAcc}})

defineFunction $ DefineFun outFun newClauses

defineDatatype $ DefineDatatype accPred ctors

function is defined using the new pattern-matching equations.

Finally, the data type is defined by its constructors and the rewritten

class resolution and a corresponding constructor of the accessibility predicate.

Because each equation is simultaneously translated into a new equation
and a rewritten definition.

Now that the accessibility predicate has been declared, the new

function, which uses it, must be declared. The new functions
arguments are those of the previous function along with the accessibility
predicate applied to the other arguments.

let accArg = MkFunArg `{{acc}}

(mkApp (Var accPred)
  (map (Var . name) args))

Explicit

NotErased

In this code, Var is a term constructor for variable references, taking a
name as an argument, and name projects the name field from
FunArg. In Idris’s surface syntax, this argument would be written
(acc : accPred a), where accPred is the name provided for the
accessibility predicate and a is the names of the previous arguments
to the function. In the case of qs, this will be (acc : QSAcc xs).

The new argument accArg is added to the signature for the rewritten
function:

declareType $

Declare outFun (args ++ [accArg]) res

We declare the rewritten function's type and the type constructor of
the accessibility predicate first, because the constructors of the
data type may need to refer to the function, and those will not be
typeable if it does not yet exist.

Now that the typing rules for the rewritten function and the
accessibility predicate have been added, it is time to compute the
constructors of the data type and the pattern-matching equations for
the rewritten definition. The input to the transformation is the list of
original clauses:

DefineFun _ oldClauses <- lookupFunDefnExact inFun'

Because each equation is simultaneously translated into a new equation
and a corresponding constructor of the accessibility predicate,
the result of processing the clauses is unzipped into constructors and
new pattern-matching equations. The processClauses function is a
direct implementation of Bove and Capretta’s algorithm.

(newClauses, ctors) <-

unzip <$> processClauses Z oldClauses

Finally, the data type is defined by its constructors and the rewritten
function is defined using the new pattern-matching equations.

defineDatatype $ DefineDatatype accPred ctors
defineFunction $ DefineFun outFun newClauses

We can invoke bc using the %runElab directive:

%runElab (bc `\{quicksort\} `\{{qs'}}) `\{{QSAcc}}}

The resulting definitions are the TT equivalents of:

data QSAcc : List Nat -> Type where

QSAccCase0 : QSAcc []
QSAccCase1 : QSAcc (filter (< x) xs) ->
  QSAcc (filter (>= x) xs) ->
  QSAcc (x :: xs)

and

qs' : (xs : List Nat) -> QSAcc xs -> List Nat
qs' [] QSAccCase0 = []
qs' (x :: xs) (QSAccCase1 l r) =
qs' (filter (< x) xs) l ++
  x :: qs' (filter (>= x) xs) r

It is now up to a user to show that QSAcc xs is inhabited for all lists
xs by defining a helper such as the following:

qAcc : (xs : List Nat) -> QSAcc xs

This particular formulation is convenient when there may be some
additional preconditions for totality. In cases where the function is
actually total, it may be convenient to instead generate an unfulfilled
proof obligation that the entire domain is accessible as well as a
wrapper with the same signature as the original function. This proof
obligation can then be discharged using all of Idris’s standard tools
for interactive development. The wrapper then uses the result of
this proof to call the rewritten function. Because this mechanism
did not require modifications to the Idris compiler, users are free to
customize it however they want, while still being able to rely on the
unmodified underlying type checker for correctness, and features
such as custom syntax rules3 can be used to provide a syntax akin
to that in Section 2.1

5.2 Proof Automation with Tactics

Previous versions of Idris supported a special-purpose tactic language
for writing proofs at an interactive shell, and user interfaces reminiscent of
dedicated proof assistants such as Coq were constructed to facilitate interactive proving. However, the special-purpose tactic language had a number of drawbacks: its extreme
simplicity meant that it was not possible to build abstractions or
non-trivial derived tactics, it was impossible to reuse Idris code in
tactics, and it introduced a large number of new keywords into the
language grammar. Elaborator reflection has subsumed this previous
tactic language, allowing Idris’s ordinary abstractions and definition forms to be used to build a richer language of tactics. Idris additionally includes a library of derived tactics that is named Pruviloj.
In addition to extended versions of the primitive tactics that do things like introducing multiple lambda bindings at once, Pruviloj includes features such as an induction tactic that generates its own eliminators for inductive families and tactics for automatically discharging proof
goals with absurd hypotheses. The eliminator-generation feature of
Pruviloj replaces a built-in feature of the Idris compiler, and we aim to
implement more of Idris in itself in the future.

To demonstrate the utility of the derived tactics in Pruviloj for
proof automation, we now present the definition of the match tactic
from Section 2.2. The tactic will perform the following steps when
given a goal of the form (x : t) -> P x:
1. Bind x under a lambda, introducing it as a local assumption
2. Perform induction on x, generating elimination principles as necessary
3. In each proof goal introduced by the induction, apply a non-
   inductive solver called auto.

   html#syntax-rules
This mush tactic uses derived tactics and tacticals such as intros, which introduces as many times as possible with automatically-generated names; try, which applies its argument but does nothing in case of failure; induction, which generates and applies an elimination principle to the appropriate value and motive, returning a list of holes corresponding to the cases for each constructor; and andThen, which runs its first argument, expecting a list of hole names, then applies its second argument in each of the holes.

The helper tactic auto performs the following steps:

1. Normalize the goal
2. If the goal is a universal quantification, introduce all quantified variables as local assumptions
3. Attempt to rewrite the goal with all equality proofs that are assumptions from induction
4. Attempt to dispatch the potentially-rewritten goal using a local hypothesis, or Idris’s built-in proof search if that fails

The only new Pruviloj tactic employed in the definition of auto is hypothesis, which attempts to solve the goal with each variable in the local scope, failing if none apply. The tactic search is defined in the Prelude as search’ 100 []. All of the control structures, including the do-notation, the recovery operator <|>, the iteration operator for_, and the mapping operator <$> are standard Idris operators that can be used with any instances of Monad, Alternative, Foldable, and Function type classes, which are analogous to their Haskell definitions.

5.3 Deriving Type Class Instances

Haskell supports deriving of instances for a collection of built-in type classes, freeing users from the burden of writing boilerplate Boolean equality comparisons and conversions to String. Idris has no similar feature. However, using elaborator reflection, we have been able to implement deriving of type classes such as Eq and Show. While deriving of Eq and Show are implemented as traditional code generators, we have also implemented a tactic that will derive decision procedures for Idris’s propositional equality for a large class of data types, given a type signature. As decidable equality typically requires a number of pattern-matching cases that is quadratic with respect to the number of constructors in the data type, this can save a great deal of code.

5.4 Interaction with Idris Type Providers

Idris’s type providers (Christiansen 2013), inspired by F# (Syme et al. 2013), allow arbitrary side effects to be performed during type checking. This lets Idris programs refer to the world in which they are found. Type providers in Idris differ from F#'s type providers, along with analogous uses of IO in Template Haskell, in a very important way: instead of using IO to do code generation, with the generated code spliced into the program prior to further type checking, Idris type providers compute a value using IO which is then included directly in the elaborated program. In particular, an Idris type provider is a term with type IO (Provider a), where Provider is an error monad isomorphic to Either String. The elaborator for the declaration syntax:

%provide (x : a) with p

executes p as an IO action. If p returns a success, then the contents of the success are unpacked and used as the definition of x, while if p returns an error message, elaboration halts and displays the message to the user. Because they are ordinary IO actions returning an ordinary Idris data type, Idris type providers manage to stay within the abstractions of the language. Idris’s full dependent types enable these provided values to be used as codes in the technique of generic programming with universes (Altenkirch and McBride 2003; Benke et al. 2003). A single computed value can be used to instantiate an entire library of code at one particular type from a predetermined part of Idris’s type system.

However, generic programming with universes does not come for free. Types described with a universe encoding do not enjoy the conveniences of native data types, such as convenient notations for pattern matching, and techniques like pattern synonyms cannot easily be specialized to a particular returned code. The syntactic overhead is substantial and difficult to avoid. Additionally, while Idris’s optimizer and partial evaluator (Brady and Hammond 2010) may be able to reduce or remove the cost of carrying around codes at run time, this cannot be relied upon in general, which may lead to unpredictable changes in performance. For these reasons, generic programming with universes is not applicable to all type providers, and code generation may sometimes be a better option.

By first using a type provider to discover information about the surrounding world and then an Elab script to generate code, the full expressiveness of an approach based on code generation becomes available, while still maintaining the possibility of using ordinary Idris IO actions for the effectful portions. Here, we demonstrate the definition of a type provider combined with an Elab script that, together, generate a record type based on a list of field/type pairs in a text file. Similar metaprograms are used for purposes such as generating data types based on database schemas.

An Idris record type consists of a single-constructor data type along with a collection of getters and setters that are named according to a particular convention. Getters project a single field from an instance of the record type, while setters construct a new instance that is the same as an old one, except with one field replaced by a new value. In Idris, the syntax record {f} r projects a field from r using the getter named f, while record {f=:y} r invokes a setter named set_f to construct a modified version of r. In other words, to generate a record type named R with fields f₁ : t₁, ..., fn : tn, we generate the following type declaration:

```
data R : Type where
    MkR : (f₁ : t₁) → ... → (fn : tn) → R
```

We also generate getters and setters for each field fi:

```
f₁ : R → t₁
f₁ (MkR a₁ ... aₙ) = a₁
set_f₁ a (MkR a₁ ... aₙ) = (MkR a ... a₁ aₙ₊₁ ... aₙ)
```

mush = Elab ()
mush =
do attack
x <- gensym "x"
intro x
try intros
  induction (Var x) 'andThen' auto
solve

hs <- map fst <$> getEnv
for_ hs $ (\h ➜ try (rewriteWith (Var h))
 hypothesis <|> search
solve

hs <- map fst <$> getEnv
for_ hs $ (\h ➜ try (rewriteWith (Var h))
 hypothesis <|> search
solve
In our example type provider, record fields can be either strings, floating-point numbers, or arbitrary-precision integers. We represent these types using the following data type:

data Ty = STRING | INTEGER | DOUBLE

We use the type `List (String, Ty)` to represent the field names and types of a particular record type. A type provider to read a description from a text file looks like an ordinary Idris program:

```idris
getRec : String -> IO (Provider (List (String, Ty)))
getRec file =
  do (Right contents) <- readFile file
     case parseStr rec contents of
        | Left err => return (Error (show err))
        | Right info => return (Provide info)
```

It attempts to read the contents of the given file, passing any errors on to the type provider error constructor. Then, it invokes a combinator parser on the contents of the specified file. Parse errors are returned as type provider errors, while successful results are returned in the `Provide` constructor to be spliced in.

The type provider can be invoked on a text file using the following syntax:

```idris
%provide (info : List (String, Ty))
  with getRec "ProviderTest.txt"
```

which results in the top-level name `info` being bound to the field specification in the text file, assuming that the file exists and is syntactically correct. The next step is to invoke an elaboration script to convert the field specification into a real Idris record type using the following syntax:

```idris
%runElab (genRecord `{{MyRecord}} info)
```

The elaboration script `genRecord`, defined in Figure 6, is responsible for generating the data type and its getters and setters. It is invoked with a name and a record specification as arguments.

The first step in `genRecord` is to place the user’s provided name into the current namespace using an helper operation called `inThisNS`, which queries `Elab` for the namespace and wraps it around the name provided as an argument. Then, it declares a zero-argument data type with the namespaced name. After that, it generates a name for the constructor using `ctorName`, which prepends `Mk` to the user-provided record type name. Next, it adds the constructor, using the helper function `fieldArg` to convert `(String, Ty)` pairs into the `FunArg` type described in Section 3.3. Finally, it adds the constructor to the data type and calls a helper operation to define the accessors according to the scheme described earlier in this section. Having done this, the type is now an ordinary Idris record type and can be used with Idris’s record syntax.

Additionally, it will be compiled just like any other Idris record type. Elaborator reflection does not presently support reusing Idris’s own record type elaborator, but this would be a useful extension in the future.

6. Related Work

Reflection in Agda (pre-2016) Agda’s elaborator is different in a number of ways from Idris’s: the core language is not strictly delineated from the high-level language and a global metavariable context is used instead of an Oleg-style development calculus (Norell 2007). Agda’s type checker can be seen as an elaborator from terms that may contain metavariables to terms that do not. In previous versions of Agda, a number of quoting and unquoting splicers were available, along with a data type representing reified high-level Agda terms and definitions. Because the splicing operator acted on high-level Agda terms, it could perform elaboration to do things like solving implicit arguments. Previous work by van der Walt and Swierstra (2012) and Kokke and Swierstra (2015) demonstrated how to use this to encode non-trivial proof automation in Agda itself, and both cite the fact that this proof automation is performed within Agda itself as major advantages. Unfortunately, both systems needed to reimplement much of the infrastructure that is part of the Agda implementation, such as unification, and each was forced to duplicate information about the types of values that Agda already has in its global context.

Reflection in Agda (post-2016) Inspired by Idris’s elaborator reflection, the Agda language has recently adopted a form of elaborator reflection in place of the previous API. Since version 2.5.1, Agda exposes its elaboration monad, including the unification and metavariable machinery, to user elaboration scripts. Because the details of its elaboration mechanism differ substantially from those of Idris, so do the operations exposed in its reflected elaborator.

Direct vs. Indirect Reflection In his Ph.D. thesis on reflection in Nuprl, Barzilay (2006) contrasts two approaches to the design of a reflection system: indirect reflection, in which the aspects of the language that are to be reflected are re-implemented internally, and direct reflection, in which the underlying implementation is invoked directly. Barzilay points out a number of advantages of the direct approach to reflection: indirect representations of terms have an exponential increase in memory usage as levels of quotation increase, it becomes more difficult to ensure the correspondence between both implementations of the system, and having two implementations of each feature vastly increases the work that is necessary to develop and maintain new language features. The previous work on reflection in Agda by van der Walt and Swierstra (2012) and Kokke and Swierstra (2015) is an excellent example of the burden that indirect reflection places on language and tool authors.

While Idris’s reflection system uses an indirect representation of reified terms, the exponential blowup is not a major problem because typical applications require no more than one level of quotation. Additionally, because the quoted terms are in Idris’s small, stable core language, concerns about ongoing development are less pressing than they would be if the high-level language were quoted. The large, complex parts of Idris, such as the type checker and the elaboration machinery, are reflected directly. For example, the elaboration strategy for Idris’s `rewrite ... in ...` syntax, which rewrites the current goal type using an equality proof, is exposed directly as the `rewriteWith` primitive, rather than requiring that it be re-implemented in Idris code. Likewise, the `apply` tactic, which has proven to be highly useful in both

---

Figure 6. Generating a record type
the elaborator that is written in Haskell and in reflected elaboration scripts, is exposed directly rather than being implemented in a library. This tactic requires somewhat complex reimplementations of the typing rules for dependent functions in order to compute the correct dependent types for the holes that the operator will be applied to, and it makes little sense not to share the implementation.

**Template Haskell** Template Haskell (Sheard and Jones 2002) is a metaprogramming system for Haskell in which metaprograms are monadic actions that transform Haskell programs. It has been enabled a great deal of interesting work, including prototype implementations of generic programming extensions (Norell and Jansson 2004). Template Haskell’s Q monad, however, does not supply effects related to the elaboration infrastructure of implementations such as GHC. Instead, it provides only the ability to produce fresh names and the ability to look up definitions from the global context.

**Tactic-Based Interactive Proof Assistants** There is a rich tradition of interactive proof assistants based on automation through tactic languages. Indeed, the ML family of programming languages began as a tactic language for the Edinburgh LCF system, and Nuprl also uses a dialect of ML for proof automation. Coq has a dedicated DSL for writing tactics, known as Ltac (Delahaye 2000), with features such as backtracking and matching against terms and contexts. Idri’s elaboration infrastructure intentionally resembles the control structures and features found in these systems. However, the reflected elaborator uses Idris as its own metalanguage, allowing code re-use across programs, specifications, and proofs automation or metaprogramming.

**MTac** MTac (Ziliani et al. 2013) is an alternative tactic language for Coq that, like Elab, provides a monadic interface to proof automation, using Coq’s own term language. In MTac, a computation with type \( \text{T} \) is a tactic that will either solve a goal of type \( \text{T} \) or fail. In other words, the type of a tactic provides information about the goals to which it can be applied. While they are superficially similar, MTac and elaborator reflection serve different purposes. MTac is a dedicated proof automation language, specifically tailored to that purpose, while elaborator reflection is intended to be a general-purpose means of language extension. Implementing MTac’s primitives would be an interesting use of the reflected elaborator.

**Internal Metaprogramming in Type Theory** Altenkirch and McBride (2003) and Benke et al. (2003) pointed out that many applications of generic programming can be internalized directly into type theory, using a universe encoding. Later work by Chapman et al. (2010) showed that primitive facilities for defining data types can be replaced by a single self-describing type of descriptions, unifying generic programming with ordinary programming. Chapman et al. refer to this technique as *levitation*. Inspiring as this work is, many practical problems remain to be solved before we can build our languages on top of levitation. In particular, we do not yet have predictable optimization strategies to eliminate all extraneous data type descriptions at run time, and implementations of techniques like levitation are notoriously slow when compared to native data types in existing type checkers. While we wait for the missing technology, other techniques are relevant and interesting.

Other, special-purpose universes have been used to write non-trivial and interesting metaprograms, including some of the ones that we have implemented with elaborator reflection, internally to type theory. In particular, McBride (2015) showed how to generically describe general recursion in Agda, including a generic version of the Bove-Capretta predicates that are implemented using an encoding based on Dybjer and Setzer’s 1999 encoding of inductive-reursive definitions. Devriese and Piessens (2013) provided a large collection of metaprogramming tools implemented directly within Agda. Because these metaprograms are written in the type theory, they are also verified, and they have all the other advantages of programs written in type theory.

While this line of work is exciting, there are many important problems to be solved before we can rely on internalized reasoning for all of our metaprogramming needs. First, described data types still have a significant performance overhead, both during type checking and execution, relative to native data types and definitions. Second, the need to account for every detail internally can lead to epic amounts of code and work to automate even simple procedures — Devriese and Piessens’s metaprogramming system required approximately 1200 lines of code to implement a generic printer, much of which is a mandatory correctness proof for the metaprogram. By contrast, programs in the reflected elaborator are similar in length and complexity to traditional tactic scripts and code generators. An implementation of type class instance derivation for Show using Idris’s elaborator reflection that handles more data types than Devriese and Piessens’s implementation and must take care of additional details like generating instance objects is less than 300 lines of code. Finally, a number of separate universes for metaprogramming exist, with different strengths and weaknesses, and code written for one is not immediately applicable to the others. There is hope on the horizon — Magalhães and Löh (2014) demonstrate a universe for Haskell datatypes from which many other commonly-derived universe encodings can themselves be derived. A similar solution might work for universes of dependent data types.

While one of these universes or tools could be given special syntactic and run-time support, privileging a single metaprogramming system for fast execution and tool support might prevent progress with others. Additionally, metaprograms written in Elab can use verified internal metaprograms’ tactics by generating the appropriate code. We need not make an all-or-nothing choice between external and internal metaprogramming, and elaborator reflection could even be used to ease some of the overhead of using internal metaprogramming features, just as Idris’s elaborator eases the burden of working in a very small language like TT.

7. Discussion

We have described an approach to metaprogramming called *elaborator reflection* that is applicable to languages that have a type checker that synthesizes terms in either a separate core language or a much more explicit version of the source language. In Idris, elaborator reflection has enabled features to be moved from the compiler to libraries, making the implementation simpler and providing users with more flexibility. It has also replaced a previous special-purpose tactic language. The features provided by the elaboration framework, such as incremental construction of terms and higher-order unification, have proven to be useful for a variety of metaprogramming tasks. Additionally, elaborator reflection provides a framework within which other metaprogramming systems can be implemented.

Idris’s reflected elaborator falls short of a complete tool for language extension. It is not yet possible to replace or extend the built-in elaboration procedures for already-existing language features, and the need to invoke elaboration scripts explicitly using \%runElab means that they are not as convenient as built-in features of Idris. We hope to lift these limitations in the future.

In elaborator reflection, metaprograms have direct access to the infrastructure of the elaborator itself, which provides precisely those tools that were already necessary to elaborate the language in question. The concept of elaborator reflection is not limited to implementations of dependent type theory. It is applicable to any language in which concepts outside of the language’s own semantics are used to implement a type checker or elaborator. While Agda has already adopted elaborator reflection, it is instructive to consider how the concept might look in other languages.
Elaborators that are implemented as one monolithic program are less likely to be useful in a reflection setting. However, we expect that realistic elaborators will contain a substantial support library, even if it does not resemble a collection of tactics. In many cases, this library will also be useful for metaprogramming within the language.

For instance, the GHC implementation of Haskell performs elaboration by arranging for a constraint solver to produce explicit evidence in a small core language. System FC. Initially, a program is traversed, generating a large collection of constraints to be solved. When these constraints are solved, each solution contributes to the final program in the core language, which can be type checked independently (Vytiniotis et al. 2011). It could be interesting to design a reflection of this system, providing Haskell code with the ability to add new constraints with novel means of creating evidence terms. This could serve as an alternative to more “heavyweight” constructions such as the type checker plugin API. Furthermore, Atkey’s recent work on the relationship between type checking, elaboration and tactics (Atkey 2015), for example, suggests one possible path to wider adoption of the concept.

7.1 Future Work

Elaboration Stages As described in Section 4.3, Idris’s elaborator reflection system suffers from subtle staging issues similar to those faced by Template Haskell. It is convenient to be able to define an elaboration script and use it in the same source file. On the other hand, the close dependence of some elaboration scripts on the order of elaboration may lead to reluctance to change the elaborator in the future, due to the risk of breaking existing metaprograms. A potential solution to these issues can be found in Flatt’s 2002 paper about the Racket (then MzScheme) module system and its treatment of macros. In this module system, the visibility of names is stratified into metalevels, with compile-time macro code being inaccessible to run-time code and vice versa. Additionally, due to the possibility of metaprograms that generate or manipulate metaprograms, there is an infinite hierarchy of stages. Within a module definition, components can be defined or imported at any stage. A similar addition to the Idris module system that classifies imports and definitions by their elaboration stage could make it possible to reliably reason about elaboration times and statically reject metaprograms that do not respect staging.

Generic Programming Unlike Haskell, Idris does not have a “native” type class deriving feature. Elab scripts can be used to implement non-trivial type class deriving, but the level of abstraction is quite low. Generic deriving mechanisms that use a uniform representation of data types, such as that described by Magalhães et al. (2010) and implemented in GHC, provide a more elegant approach that additionally allows generic programs to be written as ordinary programs, rather than through code generation. Similarly, dependent types allow non-trivial data types to be encoded generically, but the dust has not yet settled around which universe encoding of data types is best. Norell and Jansson (2004) describe an application of Template Haskell to implementing novel generic programming systems. Similarly, the Shapeless generic programming library for Scala\(^5\) uses Scala’s macro facility to implement its generics, an approach discussed by Burmako (2013). We would like to follow the lead of Norell and Jansson and the Shapeless developers and use elaborator reflection to make an implementation of a high-level generic programming system convenient to use with native data types.

Alternative Proof Languages The reflected elaborator provides many useful primitives, but exposes a very low-level interface for proof construction. Users must keep track of individual holes, ensuring that each of them is filled and solved in turn. Matching against reflected terms is highly intensional, not even respecting Idris’s definitional equality, and normalization must be specifically requested. Additionally, a single proof automation API is not equally suitable for all tasks. In future, we would like to implement higher-level proof APIs by interpreting them in Elab. We have already implemented a significant portion of Agda’s former reflection API, and we would like to explore an implementation of MTac.

Self Hosting Thus far, the only non-trivial dependently typed language implemented in itself is F\(^*\) (Swamy et al. 2016). F\(^*\) was bootstrapped by writing the first version in a subset of the language that overlaps with F\#, using the F\# compiler to compile the initial version. Rather than implementing a direct compiler in the style of Idris, F\(^*\) is first translated to either F\# or OCaml, which allowed subsequent versions of F\(^*\) to be compiled using the respective compilers. Elaborator reflection suggests an alternate route to self-hosting. Given facilities for syntactic abstraction and the ability to write terms directly in the core language, the elaborators for a high-level language like Idris could be written as macros, expanding to the core language. Then, the only primitives needed for bootstrapping would be an implementation of the core type theory, the macro system, and the basic elaboration primitives.

Elaborator reflection empowers Idris’s users to take charge of their language. In a self-hosted Idris, implemented by means of a tower of reflected elaborators with a stage-respecting module system, users will have freedom comparable to that of a Lisp user. At the same time, the stability and portability of the TT type checker provides a solid grounding for new features, ensuring that they have a sensible explanation. A reflected elaborator on top of type theory provides both freedom and security.

Acknowledgments

The first author’s work was funded by the Danish Advanced Technology Foundation (Højteknologifonden) grant 17-2010-3 and by the United States National Science Foundation grant 1540276. The second author is generously supported by EPSRC grant EP/N024222/1. We would like to thank Shayan Najd, Sam Tobin-Hochstadt, and the anonymous reviewers for their feedback on a draft of this paper, as well as Ulf Norell and Peter Sestoft for discussions during the development of elaborator reflection. The #idris channel on Freenode has also been an excellent source of feedback and discussion.

References


---

\(^5\) Available from https://github.com/milessabin/shapeless


