MASR — Meta Abstract Semantics Representation

Brian Beckman

10 Apr 2023

Contents

1 Abstract			3	
2	2.1 2.2 2.3 2.4	ASDL is Moribund	3 4 4 4	
3	Cloj	ure Solves ASDL Issues	4	
4	MAS	SR Tenets	5	
5	MAS	SR Terms and Heads	6	
6	Spec	es	8	
	6.1	intent	8	
	6.2	unit	14	
	6.3	symbol	14	
	6.4	storage_type	14	
	6.5	access	14	
	6.6	deftype	14	
	6.7	presence	14	
	6.8	abi	15	
	6.9	stmt	17	
	6.10	expr	17	
	6.11	ttype	18	
	6.12	restriction_arg	19	
		binop	19	
	6.14	logicalbinop	19	
		cmpop	19	
	6.16	integerboz	19	
	6.17	arraybound	19	
	6.18	arraystorage	19	

6.19 cast_kind	19
6.20 dimension	19
6.21 alloc_arg	19
6.22 attribute	19
6.23 attribute_arg	19
6.24 call_arg	
6.25 tbind	19
6.26 array_index	19
6.27 do_loop_head	19
6.28 case_stmt	19
6.29 type_stmt	19
6.30 enumtype	19
6.31 Implicit Terms	19
6.32 Term-Like Items	20
Change Log	20

1 Abstract

Abstract Semantics Representation (ASR) is an innovative intermediate representation¹ (IR) for multiple LCompilers.² ASR is independent of the particular programming language under compilation. Current compiler front-ends targeting ASR include LFortran³ and LPython.⁴ ASR is also agnostic to the compiler back end. Current back ends targeted *from* ASR include LLVM, x86, C, and WASM⁵

Being agnostic means that it is easy to write new compilers, both at the front end and the back end. For example, LFortran predates LPython. When the need for a Python compiler arose, only a Python front end was necessary. Within a few days, a new end-to-end compiler, LPython, was created.

LCompiler back ends are completely reusable because ASR eliminates all original language syntax from the IR, in sharp contrast to typical practice, which treat semantics as decorations on syntax trees.

In addition to being more flexible, LCompilers are faster than average because optimizers are not hampered by useless syntactic structure.

Current specifications for ASR are written in ASDL,⁶ a metalanguage similar in spirit to yacc but less rich, by design.⁷ To build an LCompiler like LFortran or LPython, the ASDL grammar for ASR is parsed and a library, asrlib, in C++ is generated. Compiler front ends call this library to transform and emit ASR trees.

ASDL has several deficiencies, and MASR, ⁸ described in this document, alleviates them. We aim to replace asrlib with MASR.

This document is pedagogical, both explaining MASR and teaching how to extend and maintain its Clojure code.

2 Issues with ASDL

2.1 ASDL is Moribund

ASDL has not progressed since originally published in 1987. We know of no other projects adopting ASDL. We should replace ASDL with a modern metalanguage that has a robust, lively ecosystem.

¹https://en.wikipedia.org/wiki/Intermediate_representation

²https://github.com/lcompilers/libasr

³https://lfortran.org/

⁴https://lpython.org/

⁵https://webassembly.org/

⁶https://en.wikipedia.org/wiki/Abstract-Type_and_Scheme-Definition_

⁷https://en.wikipedia.org/wiki/Yacc

⁸pronounced "maser;" it is a Physics pun

2.2 ASDL is Incomplete

All work with ASR is currently done with opaque binary representations and code written in C++. As usual with such, it's more time-consuming and error-prone than necessary to prototype, verify, validate, visualize, modify, and debug.

2.3 ASDL is Volatile

The ASDL for ASR changes frequently, for good reasons. However, stand-aside tools like asr-tester⁹ must chase the ASDL specification. Just keeping up with ASDL consumes almost all development time for asr-tester. We should unify the language for expressing ASR with the tools that verify and test ASR.

2.4 ASDL is Ambiguous

There are many syntactic and semantic ambiguities in the ASDL grammar. For example, the specification integer* might mean an ordered collection of integer or an unordered collection, with duplicate elements allowed or not.

ASDL cannot express such distinctions. In practice, the C++ implementations implicitly make these distinctions. In one place integer* means an ordered collection. In another place, integer* means an unordered collection, but only the generated C++ code knows where.

Hiding fine distinctions in specifications in the generated code is not good engineering practice in the face of a known better solution.

3 Clojure Solves ASDL Issues

ASR expressions, being trees, have a natural representation in S-Expressions. ¹⁰ Clojure, being a modern Lisp, natively handles S-Expressions. Clojure is modern. Clojure has a robust, lively ecosystem.

Clojure.spec, ¹¹ is a *force majeure* for precision, completeness, verification, and validation. The collection of MASR specs amounts to a meta-type system for ASR.

Clojure specs are arbitrary predicates, capable of expressing type-system logic beyond typical, hard-coded type systems. That flexibility affords new opportunities, say for experiments in dependent types and concurrency types.¹². In the short run, clojure spec will make type constraints for ASDL explicit and manifest, and will relieve the burden on C++ programmers to manage implicit constraints.

This document may lag the Clojure code. It may also lag the current state of asrlib, at least until MASR replaces asrlib. The document mirrors an ASDL snapshot.¹³

⁹https://github.com/rebcabin/asr-tester

 $^{^{10}}$ https://en.wikipedia.org/wiki/S-expression

¹¹https://clojuredocs.org/clojure.spec.alpha

¹²https://rholang.io/

 $^{^{13} \}verb|https://github.com/rebcabin/masr/blob/main/ASR_2023_APR_06_snapshot.asdl|$

4 MASR Tenets

- **Hash-Maps** ASR entities shall be hash-maps with fully-qualified keywords as keys (see Section 6.1 for motivating example).
- **Multi-Specs** ASR entities shall be recursively checked and generated via Clojure multi-specs. ¹⁴
- **Explicit** ASR entities shall contain all necessary information, even at the cost of verbosity. Defaults are not permitted.
- **Syntax Sugar** Constructor functions for ASR entities may allow default values for keyword arguments (see Section 6.11 for an example and see Issue 3 on MASR's GitHub repo.

¹⁴https://clojure.org/guides/spec

5 MASR Terms and Heads

Terms are the "objects" or "productions" of ASR, items to the left-hand side of an equals sign in the ASDL grammar. Table 1 exhibits terms that are

- explicitly specified in ASDL, like symbol or dimension
- used but not defined in ASDL, like symbol_table
- term-like but not defined in ASDL, like identifier

The definitions in Table 1 have been abbreviated and edited for presentation.

Heads are things like Local and CaseStmt that appear on the right-hand sides of terms equations in Table 1. There are of two kinds of heads:

function-like heads — have parentheses and typed parameters,
 e.g., CaseStmt (expr*, stmt*)

enum-like heads — no parentheses, e.g., Local

MASR has a Clojure spec and syntactic sugar for each head. There are about 250 heads by a recent count.

Table 1: Terms (nodes) in the ASDL grammar (things left of equals signs):

table 1. Terms (nodes) in the ASDL grantmar (things left of equals signs).					
	term	partial expansion			
1	unit	<pre>TranslationUnit(symbol_table, node*)</pre>			
2	symbol	many heads			
3	storage_type	Default Save Parameter Allocatable			
4	access	Public Private			
5	intent	Local In Out InOut			
6	deftype	Implementation Interface			
7	presence	Required Optional			
8	abi	Source LFortranModule Intrinsic			
9	stmt	many heads			
10	expr	many heads			
11	ttype	<pre>Integer(int, dimension*) </pre>			
12	restriction_arg	RestrictionArg(ident, symbol)			
13	binop	Add Sub BitRShift			
14	logicalbinop	And Or Xor NEqv Eqv			
15	cmpop	Eq NotEq Lt LtE Gt GtE			
16	integerboz	Binary Hex Octal			
17	arraybound	LBound UBound			
18	arraystorage	RowMajor ColMajor			
19	cast_kind	RealToInteger IntegerToReal			
20	dimension	(expr? start, expr? length)			
21	alloc_arg	(expr a, dimension* dims)			
22	attribute	Attribute(ident name, attr-arg* args)			
23	attribute_arg	(ident arg)			
24	call_arg	(expr? value)			
25	tbind	Bind(string lang, string name)			
26	array_index	(expr? left, expr? right, expr? step)			
27	do_loop_head	(expr? v, expr? start expr? end, expr? step)			
28	case_stmt	CaseStmt(expr*, stmt*)			
29	type_stmt	<pre>TypeStmtName(symbol, stmt*) </pre>			
30	enumtype	IntegerConsecutiveFromZero			
	implicit				
31	symbol_table	Clojure maps			
32	symtab_id	an int			
	*term-like				
0	dimensions	dimension*, via Clojure vectors or lists			
0	atoms	int float bool nat bignat			
0	identifier	by regex			
0	identifiers	identifier*, via Clojure sets			

6 Specs

The following sections

- summarize the Clojure specs for all ASR terms and heads
- pedagogically explain the architecture and approach taken in the Clojure code so that anyone may extend and maintain it.

The architecture is the remainder from several experiments. For example, defrecord and defprotocol for polymorphism were tried and discarded in favor of multi-specs. 14

The tests in <code>core_test.clj</code> exhibit many examples that pass and, more importantly, fail the specs. We also keep lightweight, load-time tests inline to the source file for the specs, <code>specs.clj</code>. The balance between inline tests and separate tests is fluid.

The best way to learn the code is to study the tests and to run them in the Clojure REPL or in the CIDER debugger in Emacs. ¹⁵

We present the terms somewhat out of the order of Table 1. First is *intent*, as it is the archetype for several enum-like terms and heads.

6.1 intent

6.1.1 Sets for Contents

An ASR *intent* is one of the symbols

Local, In, Out, InOut, ReturnVar, Unspecified.

The spec for the *contents* of an intent is simply this set of enum-like heads. Any Clojure set (e.g., in $\#\{\dots\}$ brackets) doubles as a predicate function for set membership. In the following two examples, the set appears in the function position of the usual Clojure function-call syntax (*function args**):

If a candidate member is in a set, the result of calling the set like a function is the candidate member.

```
(#{'Local 'In 'Out 'InOut 'ReturnVar 'Unspecified} 'Local)
```

Local

When the candidate element, say fubar, is not in the set, the result is nil, which does not print:

```
(#{'Local 'In 'Out 'InOut 'ReturnVar 'Unspecified} 'fubar)
```

Any predicate function can be registered as a Clojure spec. ¹¹ Therefore the spec for *intent contents* is just the set of valid members.

¹⁵https://docs.cider.mx/cider/debugging/debugger.html

6.1.2 Specs have Fully Qualified Keyword Names

The name of the spec is ::intent-enum. The double colon in ::intent-enum is shorthand. In the file specs.clj, double colon implicitly signifies that a keyword like intent-enum is in the namespace masr.specs. In other files, like core_test.clj, the same keyword is spelled :masr.specs/intent-enum.

The names of all Clojure specs must be fully qualified in namespaces.

```
(s/def ::intent-enum
  #{'Local 'In 'Out 'InOut 'ReturnVar 'Unspecified})
```

6.1.3 How to Use Specs

To check an expression like 'Local against the ::intent-enum spec, write

```
(s/valid? ::intent-enum 'Local)
;; => true
(s/valid? ::intent-enum 'fubar)
;; => false
```

To produce conforming or non-conforming (invalid) entities in other code, write

```
(s/conform ::intent-enum 'Local)
;; => Local
(s/conform ::intent-enum 'fubar)
;; => :clojure.spec.alpha/invalid
```

To generate a few conforming samples, write

```
(gen/sample (s/gen ::intent-enum) 5)
;; => (Unspecified Unspecified Out Unspecified Local)
```

or, with conformance explanation (trivial in this case):

```
(s/exercise ::intent-enum 5)
;; => ([Out Out]
;;      [ReturnVar ReturnVar]
;;      [In In]
;;      [Local Local]
;;      [ReturnVar ReturnVar])
```

Strip out the conformance information as follows:

```
(map second (s/exercise ::intent-enum 5))
;; => (In ReturnVar Out In ReturnVar)
```

s/valid?, s/conform, gen/sample, and s/exercise pertain to any Clojure specs, no matter how complex or rich.

6.1.4 The Spec that Contains the Contents

::intent-enum is just the spec for the *contents* of an intent, not for the intent itself. The spec for the intent itself is an implementation of a polymorphic Clojure multi-spec 14 , ::asr-term.

6.1.5 Multi-Specs

A multi-spec is like a tagged union in C. The multi-spec, ::asr-term, pertains to all Clojure hash-maps¹⁶ that have a tag named ::term with a value like ::intent or ::storage-type, etc. The values, if themselves fully qualified keywords, are recursively checked.

A multi-spec has three components:

defmulti¹⁷ — a polymorphic interface that declares the tag-fetcher function,
 ::term in this case. The tag-fetcher function fetches a tag's value from
 any candidate hash-map. The defmulti dispatches to a defmethod that
 matches the fetched tag value, ::intent in this case. ::term is a fully qual ified keyword of course, but all keywords double as tag-fetchers for hash maps.¹⁸

defmethod¹⁹ — individual specs, each implementing the interface; in this case, if the ::term of a hash-map matches ::intent, then the corresponding defmethod is invoked (see Section 6.1.7 below).

s/multi-spec — tying together the defmulti and, redundantly, the tag-fetcher.²⁰

6.1.6 Specs for All Terms

Start with a spec for ::term:

```
;; like ::intent, ::symbol, ::expr, ... (s/def ::term qualified-keyword?)
```

The spec says that any fully qualified keyword, like ::intent, is a MASR term. This spec leaves room for growth of MASR by adding more fully qualified keywords for more MASR types-qua-terms.

s/def stands for clojure.spec.alpha/def, the def macro in the clojure.spec.alpha namespace. The namespace is aliased to s.

Next, specify the defmulti polymorphic interface, term, (no colons) for all term specs:

```
(defmulti term ::term)
```

 $^{^{16} \}verb|https://clojuredocs.org/clojure.core/hash-map|$

¹⁷ https://clojuredocs.org/clojure.core/defmulti

 $^{^{18}}$ https://stackoverflow.com/questions/6915531

¹⁹https://clojuredocs.org/clojure.core/defmethod

²⁰Multi-specs allow re-tagging, but we do not need that level of generality.

This defmulti dispatches to a defmethod based on the results of applying the keyword-qua-function::term to a hash-map:

```
(::term {::term ::intent ...})
```

equals ::intent.

The spec is named ::term and the tag-fetcher is named ::term. They don't need to be the same. They could have different names.

6.1.7 Spec for intent

If applying ::term to a Clojure hash-map produces ::intent, the following spec, which specifies all intents, will be invoked. It ignores its argument, _:

```
(defmethod term ::intent [_]
  (s/keys :req [::term ::intent-enum]))
```

This spec states that an *intent* is a Clojure hash-map with a ::term keyword and an ::intent-enum keyword.

6.1.8 The Multi-Spec Itself: ::asr-term

s/multi-spec ties defmulti term to the tag-fetcher ::term. The multi-spec
itself is named ::asr-term:

```
;; name of the mult-spec defmulti tag fn
;; -----
(s/def ::asr-term (s/multi-spec term ::term))
```

6.1.9 Examples of Intent

The following shows a valid example:

```
(s/valid? ::asr-term
{::term ::intent,
          ::intent-enum 'Local})
```

true

Here is an invalid sample:

false

Generate a few valid samples:

6.1.10 Another asr-term: a Pattern Emerges

To define another asr-term, specify the contents and write a defmethod. The one multi-spec, ::asr-term, suffices for all.

For example, another asr-term for an enum-like is storage-type:

```
(s/def ::storage-type-enum
  #{'Default, 'Save, 'Parameter, 'Allocatable})

(defmethod term ::storage-type [_]
  (s/keys :req [::term ::storage-type-enum]))
```

All enum-like specs follow this pattern.

6.1.11 Syntax Sugar

{:::term ::intent, ::intent-enum 'Local}, a valid asr-term entity, is long and ugly. Write a short function, intent, via s/conform, explained in Section 6.1.3:

Entities have shorter expression with the sugar:

All our specs are like that: a long-form hash-map and a short-form sugar function that does a conformance check.

6.1.12 Capture the Enum-Like Pattern in a Macro

All enum-likes have a *contents* spec, a defmethod term, and a syntax-sugar function. The following macro pertains to all such enum-like multi-specs:

```
(defmacro enum-like [term, heads]
 (let [ns "masr.specs"
       tkw (keyword ns (str term))
       tke (keyword ns (str term "-enum"))
       tki (keyword ns (str "invalid-" term))]
   ` (do
       (s/def ~tke ~heads)
                                ;; the set
       (defmethod term ~tkw [_#] ;; the multi-spec
         (s/keys :req [:masr.specs/term ~tke]))
       (defn ~term [it#]
                              ;; the syntax
         (let [st# (s/conform
                    :masr.specs/asr-term
                    {:masr.specs/term ~tkw
                     ~tke it#})]
           (if (s/invalid? st#) ~tki, st#))))))
```

Use the macro like this:

```
(enum-like
intent
#{'Local 'In 'Out 'InOut 'ReturnVar 'Unspecified})
(enum-like
storage-type
#{'Default, 'Save, 'Parameter, 'Allocatable})
```

- **6.2** unit
- 6.3 symbol
- 6.3.1 **TODO** Variable
- 6.4 storage_type
- 6.5 access
- 6.6 deftype
- 6.7 presence

6.8 abi

Abi is a rich case. It is enum-like, similar to *intent* (Section 6.1), but with restrictions. Its heads include several *external-abis*:

and one *internal-abi*, specified as a Clojure set to get the membership-test functionality:

```
(def internal-abis #{'Source})
```

The *abi-enum* spec for the contents of an *abi* term is the unions of these two sets:

```
(s/def ::abi-enum
  (set/union external-abis internal-abis))
```

Specify an additional key in a conforming *abi* hash-map with a ::bool predicate:

```
(s/def ::abi-external ::bool)
```

Add a convenience function for logic:

```
(defn iff [a b]
(or (and a b)
(not (or a b))))
```

Specify the defmethod for the *abi* itself with a hand-written generator (clojure.spec is not quite strong enough to create the generator automatically):

Generate a few conforming samples:

```
(gen/sample (s/gen (s/and
                    ::asr/asr-term
                    #(= ::asr/abi (::asr/term %))))
            5)
;; => (#::asr{:term ::asr/abi,
             :abi-enum Interactive, :abi-external true}
       #::asr{:term ::asr/abi,
;;
              :abi-enum Source, :abi-external false}
;;
       #::asr{:term ::asr/abi,
;;
              :abi-enum Source, :abi-external false}
;;
       #::asr{:term ::asr/abi,
;;
              :abi-enum Source, :abi-external false}
;;
       #::asr{:term ::asr/abi,
;;
;;
              :abi-enum Interactive, :abi-external true})
```

6.8.1 Syntax Sugar

The sugar for *abi* uses Clojure destructuring^{21, 22} for keyword arguments.

Conforming examples:

```
(abi 'Source :external false)
(abi 'LFortranModule :external true)
(abi 'GFortranModule :external true)
(abi 'BindC :external true)
(abi 'Interactive :external true)
(abi 'Intrinsic :external true)
```

Non-conforming due to incorrect boolean:

```
(abi 'Source :external true)
(abi 'LFortranModule :external false)
(abi 'GFortranModule :external false)
(abi 'BindC :external false)
(abi 'Interactive :external false)
(abi 'Intrinsic :external false)
```

²¹https://clojure.org/guides/destructuring

²²https://gist.github.com/rebcabin/a3c24be3e17135f355348c834ab14141

Non-conforming due to incorrect types or structure:

```
(abi 'Source :external 42)  ;; types are not ::bool
(abi 'Source :external "foo") ;;   |
(abi 'Source :external 'foo) ;; -=-
(abi 'Source false) ;; no :external keyword
(abi 'Source true) ;;   |
(abi 'Source 42) ;;   |
(abi 'foo true) ;;   |
(abi 'foo false) ;; -=-
```

We don't show tests of incorrect arity.

Here is the implementation of the sugar, exhibiting the destructuring technique:

6.9 stmt

6.10 expr

6.11 ttype

Ttype [*sic*] has a nested multi-spec. Ttype is an archetype for all function-like heads, just as *intent* is an archetype for all enum-like heads.

```
(defmulti ttype-head ::ttype-head)
(defmethod ttype-head ::Integer [_]
  (s/keys :req [::ttype-head ::bytes-kind ::dimensions]))
(s/def ::asr-ttype-head
  (s/multi-spec ttype-head ::ttype-head))
```

```
(defmethod term ::ttype [_]
  (s/keys :req [::term ::asr-ttype-head]))
```

6.11.1 Full Form

One may always write out ttype specs in full:

6.11.2 Sugar for Integer, Real, Complex, Logical

Sugar for ttypes comes in two varieties, *light sugar* and *full sugar*. Light sugar require specs with keywords, as in:

```
(ttype (Integer- {:dimensions [], :kind 4}))
(ttype (Integer- {:kind 4, :dimensions []}))
```

Full sugar uses positional arguments, as in

```
(ttype (Integer))
 (ttype (Integer 4))
 (ttype (Integer 2 []))
 (ttype (Integer 8 [[6 60] [42]]))
```

See the tests for many examples.

- 6.11.3 TODO Character
- 6.12 restriction_arg
- 6.13 binop
- 6.14 logicalbinop
- 6.15 cmpop
- 6.16 integerboz
- 6.17 arraybound
- 6.18 arraystorage
- 6.19 cast_kind
- 6.20 dimension
- 6.21 alloc_arg
- 6.22 attribute
- 6.23 attribute_arg
- 6.24 call_arg
- **6.25** tbind
- 6.26 array_index
- 6.27 do_loop_head
- 6.28 case_stmt
- 6.29 type_stmt
- 6.30 enumtype
- 6.31 Implicit Terms

Terms used, explicitly or implicitly, but not defined in ASDL.

Some items specified in ASDL as *symbol_table* are actually *symtab_id*.

- 6.31.1 symtab_id
- 6.31.2 symbol_table
- 6.32 Term-Like Items
- 6.32.1 dimensions
- 6.32.2 atoms
- 6.32.3 identifier
- 6.32.4 identifiers

7 Change Log

2023-06-Apr :: Start.

2023-12-Apr :: enum-like specs