

RIF RuleML Rosetta Ring: Round-Tripping the Dlex Subset of Datalog RuleML and RIF-Core

*The International RuleML Symposium on
Rule Interchange and Applications
Las Vegas, Nevada, USA, Nov 5-7, 2009*

Harold Boley

Institute for Information Technology
National Research Council of Canada
Fredericton, NB, E3B 9W4, Canada

The RIF RuleML overlap area is of broad interest for Web rule interchange. Its kernel, Dlex, is defined syntactically and semantically as the common sublanguage of Datalog RuleML and RIF-Core restricted to positional arguments and non-conjunctive rule conclusions, and allowing equality plus externals in rule premises (only). Semantics-preserving mappings are defined between the Dlex subset of the RIF Presentation Syntax, RIF/XML, RuleML/XML, and the Prolog-like RuleML/POSL. These are the basis for RIF RuleML feature comparison and translation. The slightly augmented mappings can be composed into a ('Rosetta') ring for round-tripping between all pairs of Dlex representations

- RuleML moves to version 1.0, and RIF to Recommendation
- Overlap area between RIF and RuleML is of broad interest for Web rule interchange
- Focus on Dlex kernel:
 - Common sublanguage of Datalog RuleML and RIF-Core
 - restricted to positional arguments and non-conjunctive rule conclusions
 - allowing equality plus externals – built-ins as defined in RIF-DTB – in rule premises (only)

- Advance RIF RuleML development and interoperation:
Semantics-preserving mappings within Dlex
- Mappings between:
 - RIF Presentation Syntax and RIF/XML
 - RIF/XML and RuleML/XML
 - RuleML/XML and Prolog-like RuleML/POSL
- Can be composed into ('Rosetta') ring for round-tripping between all pairs of Dlex representations

RIF RuleML ...

- ... feature comparison
- ... concrete translator specification
- ... language convergence

Example: Definitions of 'even' predicate:

<http://ruleml.org/rif/dlex/RIFRuleMLRosettaEven>

Prepares identifying largest common subset of Datalog RuleML and RIF-Core, also to include “slotted” (RuleML) or “named” (RIF) arguments as well as objects/frames

- Formalize syntax and semantics of Dlex by specializing relevant definitions of RIF-BLD
- Both new RIF dialect, RIF-Dlex, and new RuleML sublanguage, Dlex RuleML
- Focus on Dlex since
 - 1 Contains Datalog, which is like relational databases with recursive views
 - 2 Adding equality plus externals \rightsquigarrow relevant to commercial rule systems
 - 3 Helps to understand and converge larger RIF and RuleML subsets
- Because RIF-Core (even RIF-BLD) does not support negation, Dlex does not either

- Enables round-tripping between all Dlex languages using compositions that chain one to four mappings
- Mappings with XML sources can be refined to XSLT translators
- Focus on mapping
 - χ_{dlex} from RIF Presentation Syntax to RIF/XML
 - ξ_{dlex} from RIF/XML to Fully Striped RuleML/XML
 - π_{dlex} from Stripe-Skipped RuleML/XML to RuleML/POSL

Introduction: Second Part – Define Mapping Ring (Cont'd)

- For **closed ring**, final (text-to-text) mapping ω_{dlex} from RuleML/POSL to RIF Presentation Syntax can omit/change separators and insert RIF Documents, Groups, etc.
- For **open ring**, 4-mapping composition $\pi_{\text{dlex}} \circ \sigma_{\text{dlex}} \circ \xi_{\text{dlex}} \circ \chi_{\text{dlex}}$ complemented by inverse composition $\chi_{\text{dlex}}^{-1} \circ \xi_{\text{dlex}}^{-1} \circ \sigma_{\text{dlex}}^{-1} \circ \pi_{\text{dlex}}^{-1}$

Definition (Alphabet)

The **alphabet** of the Dlex presentation syntax consists of

- a countably infinite set of **constant symbols** `Const`
- a countably infinite set of **variable symbols** `Var` (disjoint)
- connective symbols `And` and `-`
- the quantifier `Forall`
- the symbols `=` and `External`
- the symbols `Group` and `Document`
- the auxiliary symbols `(,)`, `<`, `>`, and `^`

Set of connective symbols, `Forall`, `=` and `External`, etc., disjoint from `Const` and `Var`. Variables written as Unicode strings preceded with symbol “?”

Constants written as `"literal"^^symospace`, where `literal` is Unicode string and `symospace` identifies symbol space

Syntax of Dlex (Cont'd)

In this definition, *base terms* are simple or positional terms, or terms of the form `External (t)`, where t is a positional term

Definition (Term)

- 1 *Constants and variables.* If $t \in \text{Const}$ or $t \in \text{Var}$ then t is a **simple term**
- 2 *Positional terms.* If $t \in \text{Const}$ and $t_1, \dots, t_n, n \geq 0$, are base terms then $t(t_1 \dots t_n)$ is a **positional term**. Positional terms correspond to usual terms and atomic formulas of classical first-order logic
- 3 *Equality terms.* $t = s$ is an **equality term**, if t and s are base terms
- 4 *Externally defined terms.* If t is a positional term then `External (t)` is an **externally defined term**. External terms are used for representing built-in functions and predicates



Syntax of Dlex (Cont'd)

Dlex distinguishes subsets of set Const of symbols, including *predicate symbols* and *function symbols*

Any term of form $p(\dots)$, where p is a predicate symbol, is also an **atomic formula**

Equality terms are also atomic formulas. An externally defined term of form $\text{External}(\varphi)$, where φ is an atomic formula, is also an atomic formula: an **externally defined** atomic formula

Definition (Formula, Condition Language)

A **formula** can have several different forms and is defined thus:

- 1 **Atomic:** If φ is an atomic formula then it is also a formula
- 2 **Condition formula:** A **condition formula** is either an atomic formula or a formula that has form of a *conjunction*:
If $\varphi_1, \dots, \varphi_n$, $n \geq 0$, are condition formulas then so is $\text{And}(\varphi_1 \dots \varphi_n)$, called a *conjunctive* formula

Definition (Formula, Rule Language)

- ③ *Rule implication*: $\varphi :- \psi$ is a formula, called *rule implication*, if:
 - φ is an atomic formula that is not an equality term,
 - ψ is a condition formula, and
 - the atomic formula in φ is not an externally defined term (i.e., a term of the form `External(...)`) and does not have such a term as an argument
- ④ *Universal rule*: If φ is a rule implication and $?V_1, \dots, ?V_n$, $n > 0$, are distinct variables then `forall ?V1 ... ?Vn (φ)` is a *universal rule* formula. It is required that all the *free* variables in φ occur among the variables $?V_1 \dots ?V_n$ in the quantification part. An occurrence of a variable $?v$ is *free* in φ if it is not inside a substring of the form $Q ?v (\psi)$ of φ , where Q is a quantifier (`forall`) and ψ is a formula. Universal rules will also be called ***Dlex rules***

Definition (Formula, Rule Language, Cont'd)

- 5 *Universal fact*: If φ is an atomic formula that is not an equality term and $?V_1, \dots, ?V_n, n > 0$, are distinct variables then $\text{Forall } ?V_1 \dots ?V_n (\varphi)$ is a *universal fact* formula, provided that all the free variables in φ occur among the variables $?V_1 \dots ?V_n$.

Universal facts are treated as rules without premises

- 6 *Group*: If $\varphi_1, \dots, \varphi_n$ are Dlex rules, universal facts, variable-free rule implications, variable-free atomic formulas, or group formulas then $\text{Group}(\varphi_1 \dots \varphi_n)$ is a *group formula*.

Group formulas are used to represent sets of rules and facts. Note that some of the φ_i 's can be group formulas themselves, which means that groups can be nested

Definition (Formula, Rule Language, Cont'd)

⑦ *Document*: An expression of form

$\text{Document} (directive_1 \dots directive_n \Gamma)$

is a *document formula*, if

- Γ is an optional group formula; it is called the group formula *associated* with the document
- $directive_1, \dots, directive_n$ is an optional sequence of *directives*. We will assume this sequence to be empty for simplicity



Semantics of Dlex: Semantic Structures

We use TV as set of semantic truth values $\{t, f\}$

Definition (Semantic structure)

A **semantic structure**, \mathcal{I} , is a tuple of the form
 $\langle TV, D, D_{ind}, D_{func}, I_C, I_V, I_F, I_=:, I_{external}, I_{truth} \rangle$

Here D is a non-empty set of elements called the **domain** of \mathcal{I} ,
and D_{ind} , D_{func} are nonempty subsets of D

D_{ind} is used to interpret elements of `Const` that occur as
individuals and D_{func} is used to interpret constants that occur
as function symbols

As before, `Const` denotes set of all constant symbols and `Var`
set of all variable symbols

Definition (Semantic structure, Cont'd)

The other components of \mathcal{I} are *total* mappings defined thus:

① I_C maps Const to D .

This mapping interprets constant symbols.

In addition, it is required that:

- If a constant, $c \in \text{Const}$, is an *individual* then it must be that $I_C(c) \in D_{\text{ind}}$
- If $c \in \text{Const}$, is a *function symbol* then it must be that $I_C(c) \in D_{\text{func}}$

② I_V maps Var to D_{ind} . Mapping interprets variable symbols

③ I_F maps D to functions $D_{\text{ind}}^* \rightarrow D$ (here D_{ind}^* is set of all finite sequences over D_{ind}). Interprets positional terms

- If $d \in D_{\text{func}}$ then $I_F(d)$ must be a function $D_{\text{ind}}^* \rightarrow D_{\text{ind}}$
- This implies that when function symbol is applied to arguments that are individual objects then result is also an individual object

Definition (Semantic structure, Cont'd)

- 4 $I_=$ is mapping of form $D_{\text{ind}} \times D_{\text{ind}} \rightarrow D$.
It gives meaning to equality operator
- 5 I_{external} is mapping that is used to give meaning to `External` terms. It maps symbols in `Const` designated as external to *fixed functions* of appropriate arity. Typically, external terms are invocations of built-in functions or calls to external definitions, and their fixed interpretations are determined by specification of those built-ins and external definitions
- 6 I_{truth} is mapping of form $D \rightarrow TV$.
It is used to define truth valuation for formulas

Definition (Semantic structure, Cont'd)

We also define generic mapping from terms to \mathbf{D} , denoted by I (the same symbol as used to denote semantic structures, but note different font)

- $I(k) = I_C(k)$, if k is a symbol in Const
- $I(?v) = I_V(?v)$, if $?v$ is a variable in Var
- $I(f(t_1 \dots t_n)) = I_F(I(f))(I(t_1), \dots, I(t_n))$
- $I(x=y) = I_=(I(x), I(y))$
- $I(\text{External}(p(s_1 \dots s_n))) = I_{\text{external}}(p)(I(s_1), \dots, I(s_n))$



Definition (Truth valuation)

Truth valuation for well-formed formulas in Dlex is determined using function $TVal_{\mathcal{I}}$:

- ① *Positional atomic formulas:*

$$TVal_{\mathcal{I}}(r(t_1 \dots t_n)) = I_{\text{truth}}(I(r(t_1 \dots t_n)))$$

- ② *Equality:* $TVal_{\mathcal{I}}(x = y) = I_{\text{truth}}(I(x = y))$

- To ensure that equality has precisely expected properties, it is required that $TVal_{\mathcal{I}}(x = y) = \mathbf{t}$ if and only if $I(x) = I(y)$

- ③ *Externally defined atomic formula:*

$$TVal_{\mathcal{I}}(\text{External}(t)) = I_{\text{truth}}(I_{\text{external}}(t))$$

- ④ *Conjunction:* $TVal_{\mathcal{I}}(\text{And}(c_1 \dots c_n)) = \mathbf{t}$ if and only if

$$TVal_{\mathcal{I}}(c_1) = \dots = TVal_{\mathcal{I}}(c_n) = \mathbf{t}.$$

Otherwise, $TVal_{\mathcal{I}}(\text{And}(c_1 \dots c_n)) = \mathbf{f}$.

Empty conjunction is a tautology: $TVal_{\mathcal{I}}(\text{And}()) = \mathbf{t}$

Definition (Truth valuation, Cont'd)

5 Quantification:

- $TVal_{\mathcal{I}}(\text{forall } ?v_1 \dots ?v_n (\varphi)) = \mathbf{t}$ if and only if for every \mathcal{I}^* , described below, $TVal_{\mathcal{I}^*}(\varphi) = \mathbf{t}$

Here \mathcal{I}^* is a semantic structure of form

$\langle TV, \mathbf{D}, \mathbf{D}_{\text{ind}}, \mathbf{D}_{\text{func}}, I_G, I_V^*, I_F, I_=: I_{\text{external}}, I_{\text{truth}} \rangle$

Exactly like I , except that mapping I_V^* , used instead of I_V .
 I_V^* defined to coincide with I_V on all variables except, possibly, on $?v_1, \dots, ?v_n$

6 Rule implication:

- $TVal_{\mathcal{I}}(\text{conclusion} :- \text{condition}) = \mathbf{t}$, if
 $TVal_{\mathcal{I}}(\text{conclusion}) = \mathbf{t}$ or $TVal_{\mathcal{I}}(\text{condition}) = \mathbf{f}$
- $TVal_{\mathcal{I}}(\text{conclusion} :- \text{condition}) = \mathbf{f}$ otherwise



Semantics of Dlex: Logical Entailment

Entailment of condition formulas can be viewed as queries to Dlex groups or documents. In definitions, symbol \models generically stands for **models** or for **entails**

Definition (Models)

A semantic structure \mathcal{I} is a **model** of a formula, φ , written as $\mathcal{I} \models \varphi$, iff $TVal_{\mathcal{I}}(\varphi) = \mathbf{t}$ □

Definition (Logical entailment)

Let φ and ψ be formulas.

We say that φ **entails** ψ , written as $\varphi \models \psi$, if and only if for every semantic structure \mathcal{I} for which both $TVal_{\mathcal{I}}(\varphi)$ and $TVal_{\mathcal{I}}(\psi)$ are defined, $\mathcal{I} \models \varphi$ implies $\mathcal{I} \models \psi$ □

Mapping of the Dlex Condition Language

Paper formalizes chain of semantics-preserving Dlex mappings χ_{dlex} , ξ_{dlex} , and π_{dlex} for the Condition Languages of Datalog RuleML and RIF-Core

A Dlex mapping μ is **semantics-preserving** if for any pair φ, ψ of *Dlex* formulas for which $\varphi \models_K \psi$ is defined, $\varphi \models_K \psi$ if and only if $\mu(\varphi) \models_L \mu(\psi)$

Here \models_K and \models_L denote logical entailment in, respectively, source language K and target language L of μ

Mapping of the Dlex Condition Language (Cont'd)

Tables below define mappings μ , each row specifying translation $\mu(\text{column}_1) = \text{column}_2$

column_2 often contains applications of mapping μ to these **metavariables**

When expression $\mu(\text{metavar})$ occurs in column_2 , it is understood as recursive application of μ to **metavar**. Result of application is substituted for expression $\mu(\text{metavar})$

The μ inverse, μ^{-1} , obtained by reading table right-to-left while replacing right-hand-side recursive μ applications with left-hand-side recursive μ^{-1} applications

Mapping of the Dlex Condition Language: From RIF/XML to RuleML/XML

XML-to-XML mapping ξ_{dlex} from RIF/XML to RuleML/XML is central to chain of mappings, comparing, and bridging between, kernels of RIF and RuleML

To reduce mapping 'distance', *Fully Striped* RuleML/XML used as target (i.e., second column) of ξ_{dlex}

Stripe-skipping via mapping σ_{dlex} can be employed to obtain more compact *Stripe-Skipped* RuleML/XML, chaining to source (i.e., first column) of mapping π_{dlex}

RIF/XML	RuleML/XML (Fully Striped)
<pre> <And> <formula>conjunct₁</formula> . . <formula>conjunct_n</formula> </And> </pre>	<pre> <And> <formula>ξ_{dlex}(conjunct₁)</formula> . . <formula>ξ_{dlex}(conjunct_n)</formula> </And> </pre>
<pre> <External> <content>atomexpr</content> </External> </pre>	ξ_{dlex} (atomexpr)
<pre> <Atom> <op>pred</op> <args ordered="yes"> argument₁ . . argument_n </args> </Atom> </pre>	<pre> <Atom> <op>ξ_{dlex}(pred)</op> <arg index="1"> ξ_{dlex}(argument₁) </arg> . . <arg index="n"> ξ_{dlex}(argument_n) </arg> </Atom> </pre>
<i>continued on next page</i>	

continued from previous page

<pre><Expr> <op>func</op> <args ordered="yes"> argument₁ . . argument_n </args> </Expr></pre>	<pre><Expr> <op>ξ_{dlex} (func) </op> <arg index="1"> ξ_{dlex} (argument₁) </arg> . . <arg index="n"> ξ_{dlex} (argument_n) </arg> </Expr></pre>
<pre><Equal> <left>left</left> <right>right</right> </Equal></pre>	<pre><Equal> <left>ξ_{dlex} (left) </left> <right>ξ_{dlex} (right) </right> </Equal></pre>
<pre><Const type="st">unicodestring</Const></pre>	<pre><Const type="st">unicodestring</Const></pre>
<pre><Var>unicodestring</Var></pre>	<pre><Var>unicodestring</Var></pre>

Mapping of the Dlex Condition Language: From RIF/XML to RuleML/XML (Cont'd)

RIF employs `External` wrappers around every built-in call. RuleML distinguishes user-defined/built-in calls via attribute. A RIF advantage: explicitness. The RuleML advantage: no need to (structurally) change syntax at all call occurrences when user-defined operator becomes 'compiled' into built-in

RIF built-ins themselves are adopted by RuleML: Unlike earlier SWRL built-ins, which write n-ary functions as (1+n)-ary relations, functional RIF built-ins remain functions

RIF: `args` role tag for argument sequence. RuleML: `arg` role tags for arguments separately – omitted via stripe-skipping.

RIF **and** RuleML methods should be permitted in RIF RuleML

RIF `Constants` use symbol spaces as values of their mandatory `type` attribute. RuleML `Individuals` and `Constants` use optional types, which can be symbol spaces

Unique inverse ξ_{dlex}^{-1} obtainable: ξ_{dlex} semantics-preserving

Mapping of the Dlex Rule Language: From RIF/XML to RuleML/XML

Paper extends chain of Dlex Condition Language mappings

χ_{dlex} , ξ_{dlex} , and π_{dlex}

to a chain of semantics-preserving Dlex Rule Language mappings for Datalog RuleML and RIF-Core

Central Condition Language mapping ξ_{dlex} from RIF/XML to RuleML/XML is extended here for rules

RIF/XML	RuleML/XML (Fully Striped)
<pre><Document> <payload>group</payload> </Document></pre>	<pre><RuleML> <performative index="1"> <Assert> <formula> $\xi_{dlex}(\mathbf{group})$ </formula> </Assert> </performative> </RuleML></pre>
<pre><Group> <sentence>clause₁</sentence> . . <sentence>clause_n</sentence> </Group></pre>	<pre><Rulebase> <formula>$\xi_{dlex}(\mathbf{clause}_1)$</formula> . . <formula>$\xi_{dlex}(\mathbf{clause}_n)$</formula> </Rulebase></pre>
<pre><Forall> <declare>variable₁</declare> . . <declare>variable_n</declare> <formula>rule</formula> </Forall></pre>	<pre><Forall> <declare>$\xi_{dlex}(\mathbf{variable}_1)$</declare> . . <declare>$\xi_{dlex}(\mathbf{variable}_n)$</declare> <formula>$\xi_{dlex}(\mathbf{rule})$</formula> </Forall></pre>
<pre><Implies> <if>condition</if> <then>conclusion</then> </Implies></pre>	<pre><Implies> <if>$\xi_{dlex}(\mathbf{condition})$</if> <then>$\xi_{dlex}(\mathbf{conclusion})$</then> </Implies></pre>

Mapping of the Dlex Rule Language: From RIF/XML to RuleML/XML (Cont'd)

The first row maps the RIF root element `Document` to the RuleML root element `RuleML`, which permits (ordered) transactions of performatives including `Asserts`

The second row maps a RIF `Group` to a RuleML `Rulebase`

While a general RIF `Document` can also contain directives for a base, prefixes, and imports, a general RuleML root can also contain `Query` and `Retract` transactions

But for our special case the unique inverse ξ_{dlex}^{-1} can be obtained, keeping ξ_{dlex} semantics-preserving

(Regarding last table row, the RuleML 0.91 upgrade to 0.95 adopts RIF role tags `<if> ... <then>`, also long contemplated in RuleML, instead of `<body> ... <head>`)

- **Dlex kernel of RIF RuleML overlap** study enables round-tripping and convergence
- Positional Dlex to be expanded for remaining features of Datalog RuleML / RIF-Core, e.g. slotted/named arguments, objects/frames, and IRIs as attributes/elements
- Since our Dlex mappings already include (interpreted) functions for external calls, study should also be extended to overlap of Hornlog RuleML and RIF-BLD using (uninterpreted) functions for Herbrand terms

- Datalog RuleML is complemented by Naf Datalog RuleML, a sublanguage with often needed Negation as failure, which is also carried through to Naf Hornlog RuleML
- Since Naf constructs are wide-spread in practice, most rule engines have implemented them (including Scoped Naf in cwm/AIR and Euler)
- So their interoperation is already a focus of RIF RuleML efforts (cf. our use case: <http://ruleml.org/WellnessRules>)

On the next level, the development of the RIF Production Rule Dialect & Reaction RuleML towards a RIF RuleML Reaction Rule Dialect will provide further opportunities for joint work between W3C, OMG, and RuleML, as needed by the industry