

# Managing Study Workflow Using the Resource Description Framework (RDF)

Armando Oliva, M.D., Semantica LLC, Fort Lauderdale, USA

## ABSTRACT

A clinical study is fundamentally a collection of activities that are performed according to protocol-specified rules. It is not unusual for a single subject to undergo hundreds of study-related activities. Managing this workflow is a formidable challenge. The investigator must ensure that all activities are conducted at the right time and in the correct sequence, sometimes skipping activities that logically need not be done. It is not surprising that errors occur.

This paper explores the use of the Resource Description Framework (RDF) and related standards to automate the management of a study workflow. It describes how protocol information can be expressed in the RDF in a computable way, such that an information system can easily identify which activities have been performed, determine which activities should be performed next, and which can be logically skipped. The use of this approach has the potential to improve how studies are conducted, resulting in better compliance and better data.

## INTRODUCTION

A clinical study is fundamentally a collection of activities that are performed according to specific rules as described in the protocol. These rules describe not only when activities are conducted, but also how they are grouped in logical study components such as Visits and Elements and Epochs. It is not unusual for a single subject to undergo hundreds of study related activities. Managing this workflow is a formidable challenge. The investigator must ensure that all activities are conducted at the right time and in the correct sequence, sometimes skipping activities that logically need not be done, and ensure that the data for each activity are properly documented. It is not surprising that errors occur, which may result in protocol violations that risk the interpretability of the study results.

The Resource Description Framework (RDF) is a standard developed by the World Wide Web Consortium (W3C) to represent data and metadata along with their meaning in a rich Linked Data environment. RDF has two distinct advantages over other standards, which make it desirable for testing:

1. The RDF model, once designed, is immediately implementable. The model is computable “out of the box”
2. RDF supports linking multiple standards together in an automated fashion, significantly simplifying implementation

Representing a study workflow in a computable format has many advantages, not the least of which is the ability to automate an analysis of how well a study was conducted, how well the protocol was followed, and the automated identification of protocol violators.

This paper explores the use of the RDF and related standards to automate the management of a study workflow. One of the related standards is SPIN (SPARQL Inference Notation), which allows rules, such as activity start rules, to be expressed computationally so that commercially available off the shelf (COTS) software such as RDF inferencing engines can evaluate the rules and determine which activities go next and which can logically be skipped.

This proof of concept is a spin-off of an ongoing project within PhUSE called Clinical Study Data in RDF.<sup>1</sup> In that project, the project team has created a mini study ontology in OWL<sup>2</sup> that is capable of representing a portion of the subject level data from a pilot study in RDF,<sup>3</sup> and from which highly standards-conforming SDTM<sup>4</sup> datasets can automatically be generated. The mini study ontology from that project is leveraged in this work.

## STRATEGY

Study activities are quite varied. The mini study ontology recognizes the following important activity subclasses. Many are taken from BRIDG<sup>5</sup>, although the definitions may have been modified to facilitate expressing them computationally in the RDF.

# PhUSE 2017

Table 1: Activity Subclasses in the Mini Study Ontology

Activity Subclass	Comment
Administrative Activity	e.g. obtain informed consent, randomization
Analysis	an activity which takes as input information about other activities (e.g. observations) and provides some sort of result. Included here are derived variables, assessments, statistical analyses
Intervention	an activity which intervenes/interrupts the subject's normal routine for a specific purpose, e.g. surgery, physical therapy, substance administration. An <b>observation</b> is an intervention since the observer interrupts the subject's normal daily routine in order to make the observation.
Observation	A type of Intervention where the main purpose to measure the physical, physiological, or psychological state of an individual
Planned Activity	an activity that is planned for execution on a subject based on the protocol. A planned activity is associated with a subject and a start rule
Performed Activity	an activity that has a documented outcome (e.g.) a result
Enabled Activity	a planned activity whose start rule has been met (Rule Outcome = TRUE) and has no documented activity outcome (no result). It is ready for execution.
Prerequisite Activity	An activity that must take place (and often have a documented outcome) before another activity can take place.

A typical start rule for Planned Activity T (the “target” activity) may say:

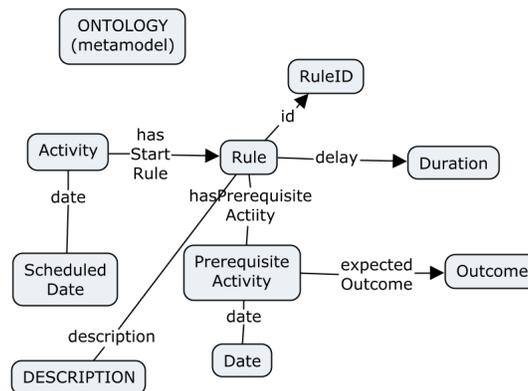
*Begin when Activity P (the “prerequisite” activity) is complete and has a specific prerequisite outcome (PO).*

The rule may also define other constraints, such as a maximum time interval following completion of P past which T should not be done, or a delay interval that must elapse following the completion of P before T can begin. Figure 1 illustrates the typical classes and relationships of a simple start rule. As an example, a screening activity may be an RPR Test for syphilis. The rule may say:

*Begin RPR Test (T) when Informed Consent Activity (P) is complete and the outcome (PO) is Informed Consent-Granted*

The rule can be expressed in the RDF using SPIN. An RDF inferencing engine can evaluate the rule as TRUE (rule condition met) or FALSE (not met), or NOT-APPLICABLE (see Rules section below). If the rule outcome is TRUE then T becomes an enabled activity, which makes it eligible for execution. Figure 1 illustrates these classes and relationships.

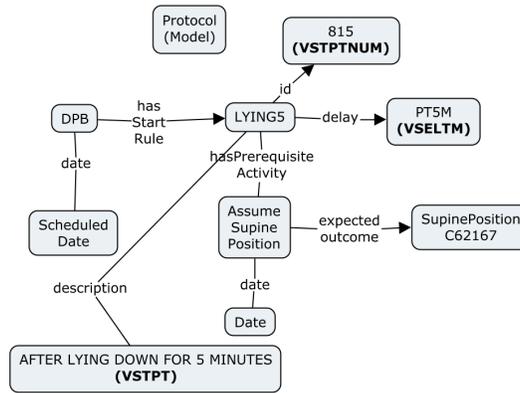
Figure 1—Simple Study Rule Ontology



In another example, consider the activity of measuring a Blood Pressure in various body positions, which is often used to detect orthostatic hypotension. The target activity T is a Blood Pressure Measurement. The prerequisite activity P is a Change In Body Position, with an expected Outcome (PO) of Supine. The rule defines a delay interval of 5 minutes, meaning T must not begin until 5 minutes after PO is achieved. The protocol-specified model is shown in Figure 2. Also shown are mappings to the SDTM Vital Signs (VS) domain (bold text in parentheses), and also a reference to CDISC Terminology for the expected PO.

# PhUSE 2017

Figure 2 -- Diastolic Blood Pressure Start Rule



Another example is a questionnaire. Each question can be considered a separate activity. Occasionally some questions are skipped based on the answer to other questions. The target question (T) “When was your last menstrual period?” may be logically skipped if the answer to the Sex question (P) was Male (PO).

## METHOD

### STUDY DESIGN

To test the strategy, a simple hypothetical study was designed. The mini study ontology was extended to accommodate activity rules and their processing. Pilot data was generated in RDF.

The pilot study has a Screening Epoch consisting of these activities:

1. Obtain Informed Consent
2. Document Demographic (i.e. Sex) Information
3. Obtain Pregnancy Test on Female Subjects Only
4. Perform RPR Test for Syphilis on all Subjects

And a Randomization Epoch consisting of one activity:

5. Randomization

Obtain Informed Consent and Randomization are each modeled as an Administrative Activity (see Table 1). The remaining three activities are each modeled as an Observation.

### RULES

Any Activity that takes as input information about other activities and generates as outcome a result is considered an Analysis in the ontology; a subclass of Activity. Rules fall in this category. A Rule is therefore modeled as a subclass of Analysis. For example:

*Sex must be Female*

is a rule that looks at the outcome of the “Document Demographic (Sex) Information” activity, the prerequisite activity, and returns a value of TRUE if the prerequisite activity outcome is “F.”

Rules may be simple or composite. A simple rule looks at information about one prerequisite activity. A composite rule looks at information about two or more prerequisite activities.

Some simple rules are met if the prerequisite activity is simply completed. These are named in the ontology using this convention: Rule\_PRCO\_(x), (“prerequisite completed”). Other rules require that a particular prerequisite activity outcome be documented (Rule\_PROUT\_(x), “prerequisite outcome” documented). A third category is a rule that is met simply when the prerequisite activity is initiated (Rule\_PRST\_(x), Prerequisite Activity = Started). Each Rule, when evaluated, may have an outcome of TRUE (Rule met, activity may proceed) or FALSE (Rule not met; activity may not proceed), or NOT-APPLICABLE (Rule not met, but the target activity can be logically skipped, e.g. Pregnancy Test in males). This gives rise to a fourth category: a Rule that resolves to TRUE when the Prerequisite Activity Outcome is met, and NOT-APPLICABLE if not met (Rule\_PROUTNA\_(x)). Each is modeled as a sub-class of Rule. The different categories of simple rules are shown in the table below.

## PhUSE 2017

Table 2: Four Categories of Simple Rules

Category	Description
Prerequisite Activity is Completed (PRCO)	Activity may begin when the prerequisite activity is complete.
Prerequisite Activity has Started (PRST)	Activity may begin when the prerequisite activity is started.
Prerequisite Activity Outcome (PROUT)	Activity may begin when the prerequisite activity outcome matches the expected outcome. If the prerequisite activity has been logically skipped (see below) then the Activity may begin.
Prerequisite Activity Outcome, Not Applicable Option (PROUTNA)	Activity may begin when the prerequisite activity matches the expected outcome, but the Activity may be ignored or skipped if the expected and actual outcomes do not match.

Finally, there is the default rule which automatically has a value of TRUE. Any activity with a start rule of the default rule can automatically begin.

Composite Rules contain two or more simple rules as subrules. The algorithm for resolving a composite rule is as follows.

1. Any subrule with a prerequisite activity that was logically skipped has its outcome set to NOT-APPLICABLE.
2. Therefore each subrule can have one of 4 possible outcomes: TRUE, FALSE, NOT-APPLICABLE, or <null>.
3. A subrule with an outcome of <null> means the activity has not yet started (and is expected to be completed). The composite rule cannot be resolved and the outcome remains <null>
4. If any of the subrule outcomes is FALSE, then the entire composite subrule outcome is set to FALSE. All subrules must be met or (logically ignored) for the composite rule to resolve to TRUE
5. If neither #4 or #5 applies, then only TRUE and NOT-APPLICABLE subrule outcomes remain and the composite rule outcome is set to TRUE.

There is a special case of a composite start rule, and that is one that determines when the Randomization Activity can begin. This start rule is in fact the eligibility assessment. This paradigm can be used to automate eligibility assessments and identification of screen failures.

The following start rules are specified in the hypothetical protocol:

Table 3 – Protocol-Specified Start Rules

Rules the depend on a specific Prerequisite Activity Outcome: Rule_PROUT	
Rule	Description
Rule_PROUT_IC:	Informed Consent Outcome = Granted
Rule_PROUT_Pregnancy:	Obtain Pregnancy Test Outcome = "NEG" (negative)
Rule_PROUT_RPR:	Perform RPR Test Outcome = "NEG" (negative)

Rules that depend on a specific Prerequisite Activity Outcome/Non-applicable option: Rule_PROUTNA	
Rule	Description
Rule_PROUTNA_Sex	Document Demographic (Sex) Information = "F" (female)

Composite Rule	
Rule	Description
Rulecomp_PROUT_Screening Observations	Determines that all Screening activities have the expected outcome (meet eligibility criteria)

For the pilot study, one can associate each activity with the following start rules.

- |   |                                      |
|---|--------------------------------------|
| 1. Obtain Informed Consent                | Default Start Rule                   |
| 2. Document Demographic (Sex) Information | Rule_PROUT_IC                        |
| 3. Obtain Pregnancy Test                  | Rule_PROUTNA_Sex                     |
| 4. Obtain RPR Test                        | Rule_PROUT_IC                        |
| 5. Randomization                          | Rulecomp_PROUT_ScreeningObservations |

## PhUSE 2017

One can also define rules for this study that depend on the start of a prerequisite activity (Rule\_PRST), but these rules are not needed for this example. These are useful for describing the start/stop of Activity groupings, e.g. Visits, Epochs, etc.

The rules are expressed using SPIN so that inferencing can automatically assign the rule outcome. For example, the SPIN Rule for all PROUT rules (rules that check for a specific outcome in the prerequisite activity) is:

```
spin:rule
# Checks to see if the rule is satisfied with the correct prerequisite activity outcome.
CONSTRUCT {
  ?this term:outcome ?ruleoutcome .
}
WHERE {
  ?this term:prerequisite ?practivity .
  ?this term:prerequisiteOutcome ?prout .
  ?practivity term:outcome ?actualout .
  BIND (IF((?prout = ?actualout), term:RuleOutcome_TRUE, term:RuleOutcome_FALSE) AS
?ruleoutcome) .
}
```

The rule being evaluated is represented as the variable `?this`. If the rule is met, a new triple is created to document the rule outcome. The rule looks for the prerequisite activity in the graph (`?practivity`) and identifies the expected outcome (`?prout`). It then compares the expected outcome (`?prout`) with the actual outcome (`?actualout`). If there is a match then the rule is satisfied and the rule outcome is set to TRUE; otherwise it is set to FALSE.

Expressing a composite rule in SPIN is more challenging because the logic is more complex. First we create one SPIN rule that identifies a Logically Skipped Activity. Programmatically this is any Activity that is a Prerequisite Activity in some Rule AND whose start rule outcome is already set to NOT APPLICABLE (e.g. Pregnancy Test in male has start rule "Sex must be Female" so its start rule outcome is already set NOT-APPLICABLE. It makes sense to set the Pregnancy Test start rule outcome to NOT-APPLICABLE as well since that test has been skipped and no outcome is expected.

The second SPIN rule counts the number of subrules that have an outcome of FALSE and the number of subrules that have a <null> outcome. If any subrule has a <null> outcome, then a prerequisite activity is not yet complete and the composite rule cannot be resolved. If any subrule has a FALSE outcome, then the composite rule outcome is FALSE. Otherwise it's TRUE since there are only TRUE or NOT-APPLICABLE subrule outcomes remaining.

### WORKFLOW MANAGEMENT

With this ontology in place, one can now populate a knowledgebase with instance data. One can envision a notional Electronic Data Capture (EDC) study workflow management application using an RDF inferencing engine that can continuously monitor the subject data as they come in and evaluates the start rule for each activity. When the start rule evaluates as TRUE, a separate SPIN rule identifies the target activity as an Enabled Activity, which means it can be scheduled for execution, according to any predefined lag interval described by the rule.

For example, assume Person\_1 is female and has no subject-level data in the database. The only Enabled Activity is Obtain Informed Consent by default. Once Informed Consent is granted and this outcome is entered in the knowledgebase, Start Rule\_PROUT\_IC resolves to TRUE, which makes both the Document Demographic (Sex) Information and Obtain RPR Test as Enabled Activities. Once the sex is entered in the knowledgebase, Start Rule\_PROUTNA\_Sex resolves to TRUE and enables execution of the Obtain Pregnancy Test Activity.

Person\_5 is male. Once informed consent is granted, and the Sex is entered as M, the Pregnancy Test rule executes but because the expected prerequisite activity outcome (Sex=F) is not met, the rule outcome is set to NOT APPLICABLE. A separate SPIN rule sets the target activity status as NOT APPLICABLE and the Pregnancy Test is not performed and not considered a protocol violation in other analyses. What happens when the next activity, Randomization, depends on the outcome of an activity whose status is NOT APPLICABLE? It is ignored but Randomization cannot begin yet since the other subrule (for the RPR Test) is <null>, not yet complete.

As one adds each activity outcome in succession to the knowledgebase, the inferencing engine correctly identifies the next activity for execution by constructing a new triple:

```
Activity_n rdf:type smm:EnabledActivity .
```

Using the test study data containing data on 8 hypothetical subjects, the ontology correctly identified the Enabled Activities based on these data at a single snapshot in time during study conduct.

# PhUSE 2017

Table 4: Study Raw Data

Subject	Informed Consent	Sex	RPR	Pregnancy
Person_1	GRANTED	F	NEG	NEG
Person_2	GRANTED	F	NEG	POS
Person_3	GRANTED	F	POS	NEG
Person_4	GRANTED	F	NEG	<null>
Person_5	GRANTED	M	<null>	N/A
Person_6	GRANTED	M	<null>	N/A
Person_7	GRANTED	M	POS	N/A
Person_8	GRANTED	M	NEG	N/A

The Enabled Activities based on these data are:

- Pregnancy Test – Subject 4
- RPR Test – Subjects 5, 6
- Randomization – Subjects 1,8

The Pregnancy Test for Subjects 5-8 (all male) were all correctly identified as NOT APPLICABLE.

## CONCLUSION

This paper describes how the RDF can be used to manage a study workflow. The pilot study, containing just a handful of activities, successfully demonstrates the validity of this approach when the study data and rules are expressed in RDF. Additional work is needed to fully understand the benefits of this approach as described as next steps below. A separate PhUSE project will demonstrate how to generate submission quality SDTM datasets from the RDF.<sup>6</sup>

This approach can also be used to automate study eligibility assessment. Eligibility criteria can be considered a composite start rule for the allocation or randomization activity (T). The Randomization Start Rule examines all the screening activities (P) and their outcomes (PO) to determine if the subject can proceed to the Randomization activity. The approach can also be used for adaptive trial designs, where the Enabled Activity may differ depending on the outcome of a previous activity, and studies where treatment is delayed until a “trigger” (e.g. headache observation) occurs.

This paradigm unlocks the information in a protocol document in a way that makes it computable so information systems can “understand” a study workflow. It has the potential to improve how studies are conducted, resulting in better compliance, fewer protocol violations, and better data.

## NEXT STEPS

A few known issues arose during the course of this pilot. These are described below and will be addressed in future iterations of the ontology. Additional steps will include supporting new requirements, as described below.

1. It is not sufficient to add new triples to the knowledgebase, but it is necessary to edit existing triples. For example, activity status (not started, ongoing, aborted, completed, not applicable) needs to be updated as activities evolve during a study. SPARQL has editing capabilities that can be leveraged.
2. Activities start rule often involves study day, e.g. Perform T x number of days after a reference start date (which is defined by a Prerequisite Activity such as Randomization or first day of dosing). The ontology currently supports these types of rules (P = Randomization, delay = study day) but more explicit support for study day can be considered.
3. The current set of rules assumes that outcomes are all categorical results. Clearly that is not the case in reality. The rules need to be able to support numeric results.
4. Piloting additional studies would be useful, for example studies with delayed start of treatment, studies that depend on the onset of a symptom or a sign (e.g. headache trials), and studies with adaptive trial designs.
5. It would be useful to add support for rules that group activities into various activity groups that are useful for study conduct and analysis, e.g. Screening Epoch begins when informed consent is initiated, and it ends when all protocol-specified screening activities are completed. These rules can help automate which activities belong in which activity groups.
6. Once a planned activity (i.e. an activity associated with a subject and a start rule) becomes an enabled activity (ready for execution) it would be desirable if the ontology could enable the generation of a proposed scheduled activity date/time. For example, if the rule states “Target activity can begin 7-14 days after

## PhUSE 2017

Prerequisite Activity is complete, the system could generate the Target activity scheduled date/time window: e.g. November 12-19, 2017. When the activity is performed, the date/time stamp can be compared to the proposed scheduled activity date/time to look for protocol violations.

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name	Armando Oliva, M.D.
Company	Semantica, LLC
Address	2509 NE 22 <sup>nd</sup> Terrace Fort Lauderdale, FL 33305, USA
Work Phone:	+1 954 489 8645
Fax:	+1 202 330 4554
Email:	aoliva@semanticallc.com
Web:	<a href="http://www.linkedin.com/in/aolivamd">http://www.linkedin.com/in/aolivamd</a> <a href="http://www.semanticallc.com">http://www.semanticallc.com</a> <a href="http://aolivamd.blogspot.com">http://aolivamd.blogspot.com</a>

Brand and product names are trademarks of their respective companies.

---

<sup>1</sup> See [http://www.phusewiki.org/wiki/index.php?title=Clinical\\_Trials\\_Data\\_as\\_RDF](http://www.phusewiki.org/wiki/index.php?title=Clinical_Trials_Data_as_RDF)

<sup>2</sup> Web Ontology Language, another standard related to RDF. See <https://www.w3.org/OWL/>.

<sup>3</sup> See <https://github.com/phuse-org/CTDasRDF>. The mini study ontology is available at <https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/study.ttl>

<sup>4</sup> Study Data Tabulation Model, see <http://www.cdisc.org/sdtm>

<sup>5</sup> The Biomedical Reference Information Domain Group, see <https://bridgmodel.nci.nih.gov/about-bridg>

<sup>6</sup> See endnote [1]