



# SMART Protocols: Summary of activities

March 2, 2017

Olga Ximena Giraldo Pasmin  
[ogiraldo@fi.upm.es](mailto:ogiraldo@fi.upm.es)  
[@olgaxgiraldo](https://twitter.com/olgaxgiraldo)

# Annotation sessions



Yhenzy Zuluaga  
Universidad del Valle  
60 protocols



Alba Rocio Corrales  
Universidad Santiago de Cali  
38 protocols



Thaura Ghneim Herrera  
Universidad ICESI  
???

2016  
(35 annotated protocols)



Jan 2017  
(77 annotated protocols)



**Goal**  
(100 annotated protocols)

**[Abstract]** The protocol describes the procedure of total RNA isolation from cells of the cyanobacterium *Synechocystis* sp. PCC 6803. This protocol is also applicable to *Synechococcus elongatus* PCC 7942 and PCC 6301, *Thermosynechococcus vulcanus*, and other unicellular and filamentous species of cyanobacteria that do not have thick polysaccharide-containing outer layers. For the latter, Trizol-containing protocols should be adapted. The yield of RNA depends on optical density of cyanobacterial culture and may reach up to 10-20 µg of total RNA per 1 ml of cell culture. RNA isolated by this method can be used for Northern blot hybridization, RT-qPCR, microarrays and Next Generation Sequencing.

## Materials and Reagents

1. Guanidinium thiocyanate (Thermo Fisher Scientific, catalog number: BP221)
2. NaCitrate
3. N-Lauroylsarcosine (Sigma-Aldrich, catalog number: L9150)
4. β-mercaptoethanol (β-ME)
5. Glacial HAc
6. NaOH

## Equipment

1. Vortex Mixer
2. Beckman high speed centrifuge with JS13.1 rotor or the like
3. 15-ml Corning centrifuge tube

Add new terms to our ontology

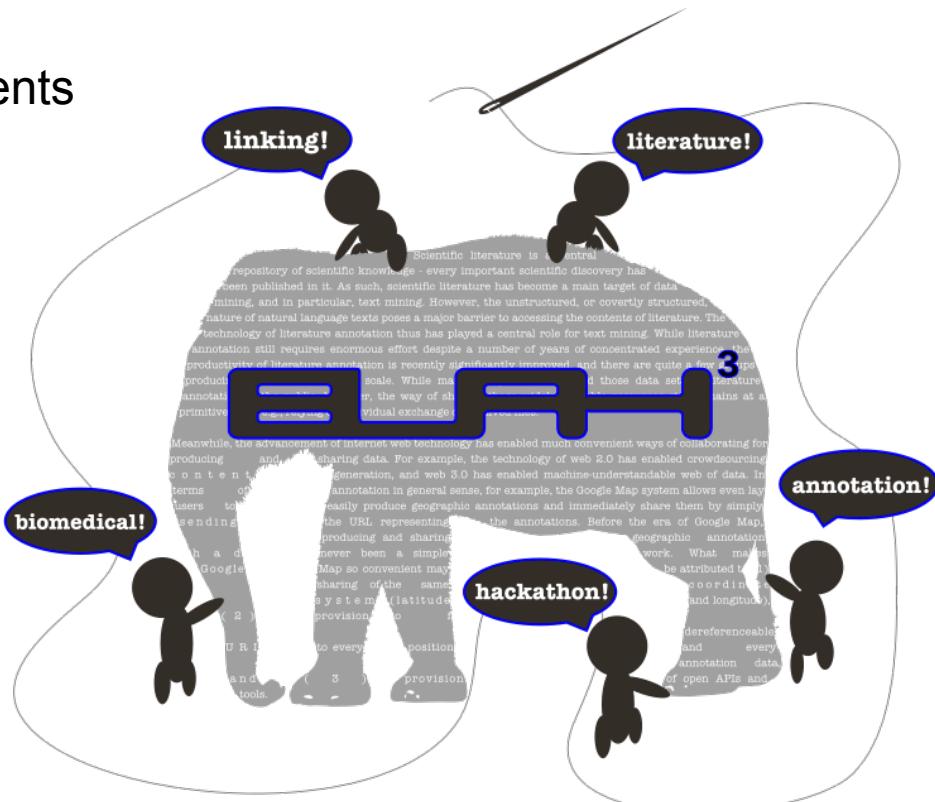
Add new synonyms to ontology terms

Terminology not curated

# Biomedical Linked Annotation Hackathon 3 (BLAH3)

## A hackathon where people

- Discuss the annotation of biomedical literature and bridging sequences-literature-resources in the web, (symposium - 1<sup>st</sup> day)
  - Work together on specific complementary efforts (hackathon – 4 days)
  - Discuss about the objectives achievements and the next steps



- Design and implement the linked data model
- Consensus about “**what is a protocol as an executable element**”



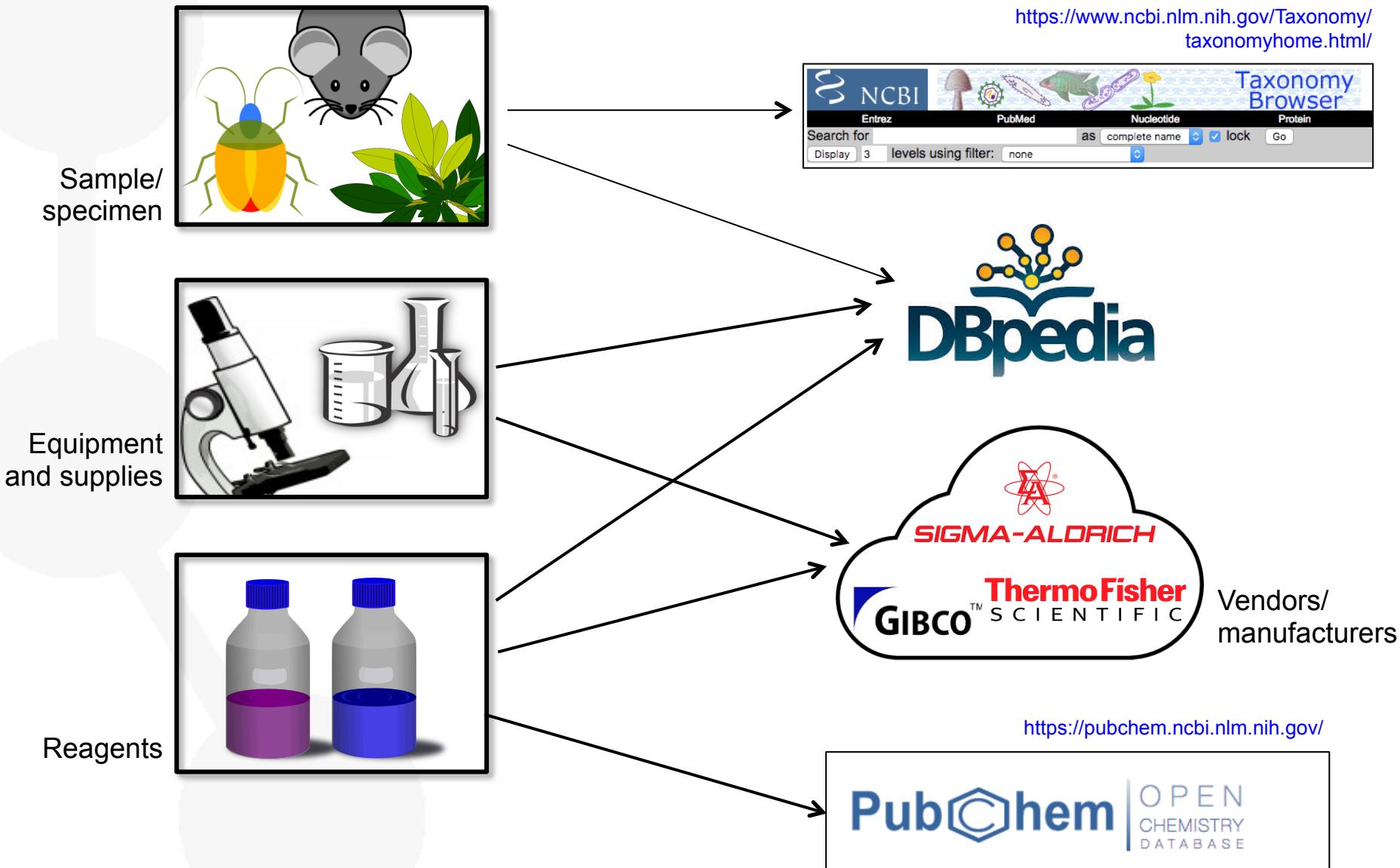
An experimental protocol is a set of instructions (or processes or steps) that should be executed following an order.

- Linkable elements
- Workflow models
- Reusable vocabularies
- Define use cases and user stories



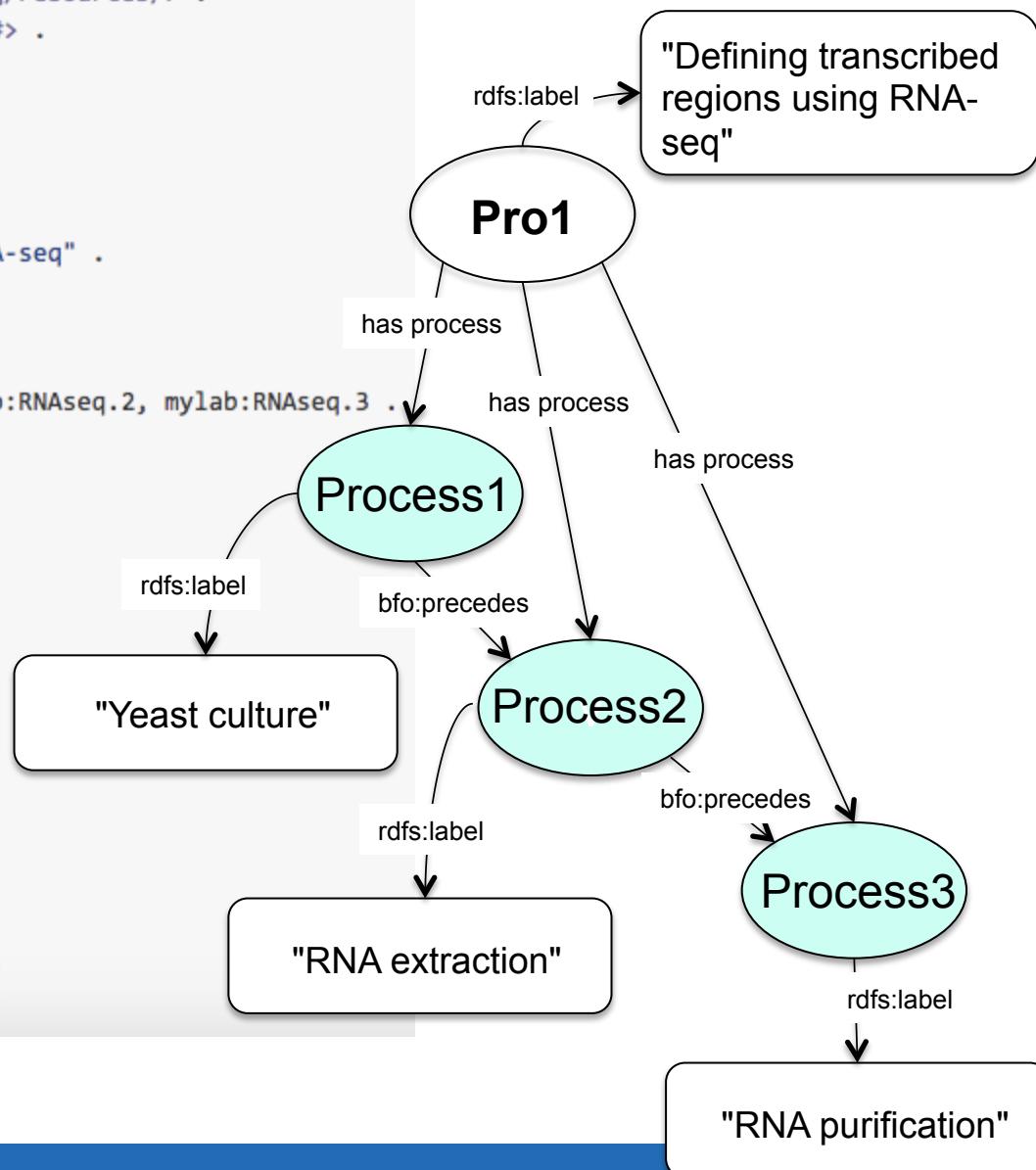
[https://www.youtube.com/watch?v=45tikBDTK\\_s](https://www.youtube.com/watch?v=45tikBDTK_s)

# linkable elements and resources



# Linked Data model

```
@prefix proto: <https://ontologies-to-be-defined.org/resources/> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
@prefix mylab: <https://mylab.org/resources/> .  
  
# RNA-seq protocol is an instance of Protocol class  
  
mylab:RNaseq a proto:Protocol ;  
    rdfs:label "Defining transcribed regions using RNA-seq" .  
  
# RNA-seq protocol has three processes  
  
mylab:RNaseq proto:has_process mylab:RNaseq.1, mylab:RNaseq.2, mylab:RNaseq.3 .  
  
# Define three processes  
  
mylab:RNaseq.1 a proto:Process ;  
    rdfs:label "Yeast culture" .  
  
mylab:RNaseq.2 a proto:Process ;  
    rdfs:label "RNA extraction" .  
  
mylab:RNaseq.3 a proto:Process ;  
    rdfs:label "RNA purification" .  
  
# Order of the execution  
  
mylab:RNaseq.1 proto:is_followed_by mylab:RNaseq.2 .  
mylab:RNaseq.2 proto:is_followed_by mylab:RNaseq.3 .
```



# More about our LD model

```
# Yeast culture process has three steps
```

```
mylab:RNAseq.1 proto:has_process mylab:RNAseq1.1, mylab:RNAseq1.2, mylab:RNAseq1.3 .
```

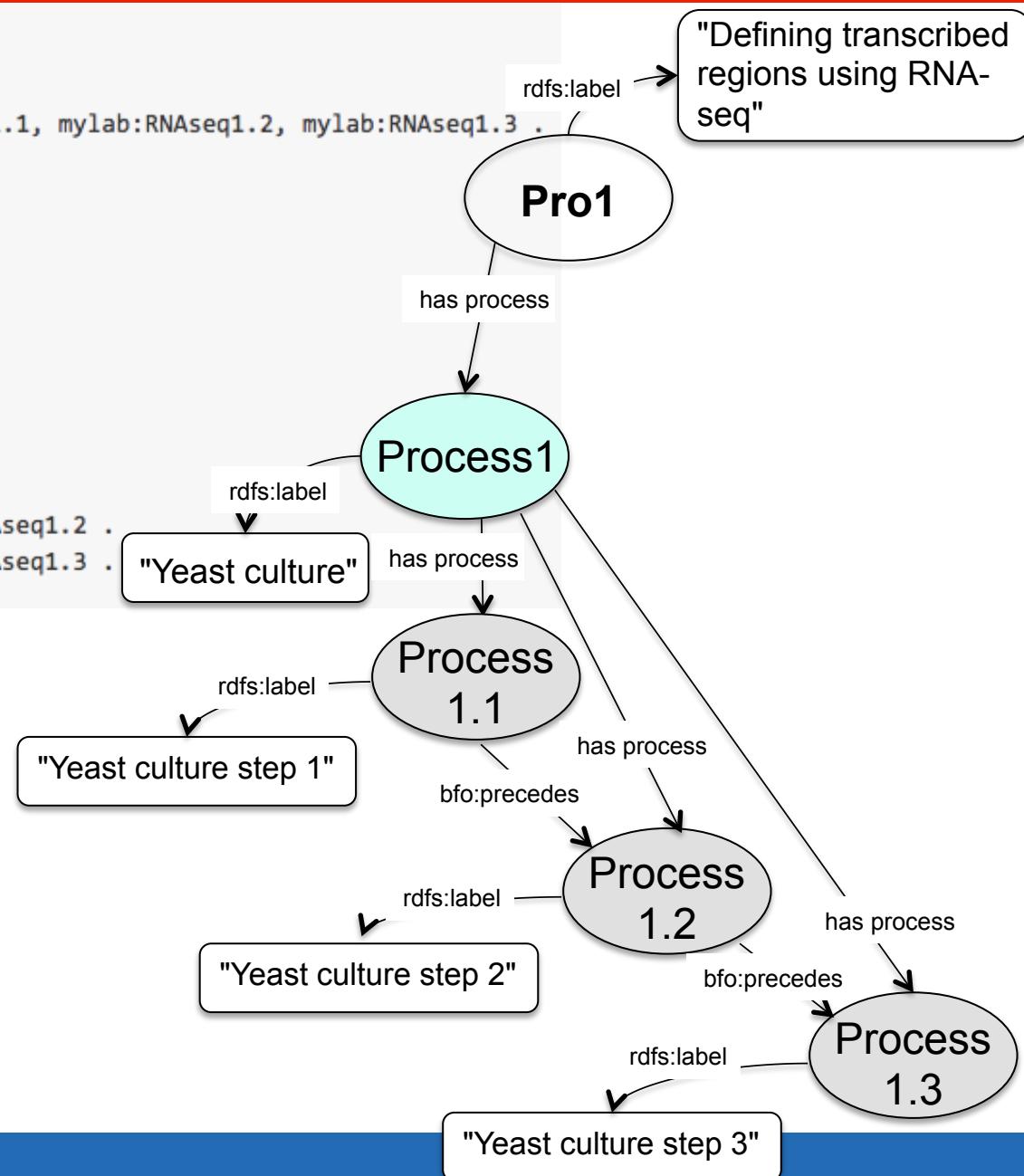
```
mylab:RNAseq1.1 a proto:Process ;  
rdfs:label "Yeast culture step 1" .
```

```
mylab:RNAseq1.2 a proto:Process ;  
rdfs:label "Yeast culture step 2" .
```

```
mylab:RNAseq1.3 a proto:Process ;  
rdfs:label "Yeast culture step 3" .
```

```
mylab:RNAseq1.1 proto:is_followed_by mylab:RNAseq1.2 .
```

```
mylab:RNAseq1.2 proto:is_followed_by mylab:RNAseq1.3 .
```



# Conditionals

```
# Repeat process1 3 times: three identical procedures  
mylab:RNAseq1.1 a proto:Process ;  
  rdfs:label "Process 1 Repeat 1" ;  
  proto:has_input mylab:RNAseq1.0-1 ;  
  proto:has_output mylab:RNAseq1.1-1 .
```

```
mylab:RNAseq1.2 a proto:Process ;  
  rdfs:label "Process 1 Repeat 2" ;  
  proto:has_input mylab:RNAseq1.1-1 ;  
  proto:has_output mylab:RNAseq1.2-1 .
```

```
mylab:RNAseq1.3 a proto:Process ;  
  rdfs:label "Process 1 Repeat 1" ;  
  proto:has_input mylab:RNAseq1.2-1 ;  
  proto:has_output mylab:RNAseq1.3-1 .
```

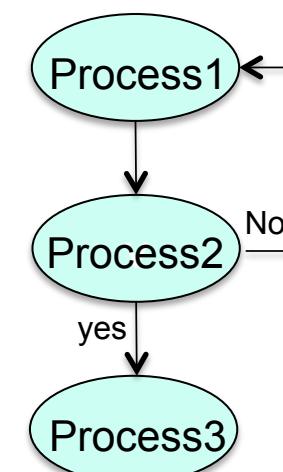
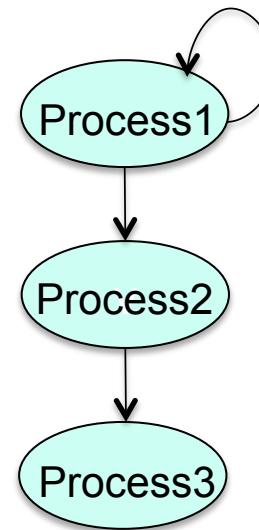
```
# two processes  
mylab:RNAseq1 a proto:Process .  
mylab:RNAseq2 a proto:Process .  
  
# process 1 is followed by process 2  
mylab:RNAseq1 proto:is_followed_by mylab:RNAseq2 .
```

```
# process 1 has a condition, and the condition has a link to the next process to be processed if the condition was f  
mylab:RNAseq1 proto:has_condition mylab:RNAseq1Cond ;  
mylab:RNAseq1Cond proto:process mylab:RNAseq2 .
```

```
# In this case, condition is just an evaluation of exit status.  
mylab:RNAseq1Cond rdfs:label "Is mylab:RNAseq1 finished correctly?"
```

Usually, the second process should be executed if the previous one is processed correctly. So the condition can be omitted, and processes are simply described as follows.

```
# two processes and a link between them  
mylab:RNAseq1 a proto:Process .  
mylab:RNAseq2 a proto:Process .  
mylab:RNAseq1 proto:is_followed_by mylab:RNAseq2 .
```



# Sparql queries

## Protocolos convertidos a Linked Data

Mouse Retinal Whole Mounts and Quantification of Vasculature Protocol [Bio-protocol Spreadsheet](#)

Scratch Wound Healing Assay [Bio-protocol Spreadsheet](#)

Soft-Agar colony Formation Assay [Bio-protocol Spreadsheet](#)

[Bio101] Subcutaneous Injection of Tumor Cells [Bio-protocol Spreadsheet](#)

Isolation of Lung Infiltrating Cell in Mice [Bio-protocol Spreadsheet](#)

Extraction of total RNA from fresh/frozen tissue (FT) [Biotechnique Spreadsheet](#)

<http://bit.do/googledocs-sparql-sp>

<https://smartprotocols.github.io/queries/>

### Retrieve the protocols using mouse as a sample

```
PREFIX sp: <http://purl.org/net/SMARTprotocol#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ro: <http://http://www.obofoundry.org/ro/ro.owl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>
```

```
SELECT ?title ?externalUri
WHERE {
    ?protocol sp:hasTitle ?title_uri .
    ?title_uri rdf:value ?title .
    ?protocol ro:hasPart ?materials .
    ?materials a sp:MaterialsSection .
    ?materials ro:hasPart ?reagents .
    ?reagents a sp:SpecimenList .
    ?reagents ro:hasPart ?reagent .
    ?reagent owl:sameAs ?externalUri .
    ?reagent sp:hasName ?nameUri .
    ?nameUri rdf:value "mouse" .
}
```

	title	name	desc	externalUri
1	"[Bio101] Subcutaneous In-jec-tion of Tumor Cells"	"CB17 scid/scid mice"	"A mouse (plural: mice) is a small rodent characteristically having a pointed snout, small rounded ears, a body-length scaly tail and a high breeding rate. The best known mouse species is the common house mouse ( <i>Mus musculus</i> ). It is also a popular pet. In some places, certain kinds of field mice are locally common. They are known to invade homes for food and shelter. Primarily nocturnal animals, mice compensate for their poor eyesight with a keen sense of hearing, and rely especially on their sense of smell to locate food and avoid predators."@en	<a href="http://dbpedia.org/resource/Mouse">http://dbpedia.org/resource/Mouse</a>
2	"Isolation of Lung Infiltrating Cell in Mice"	"Mice"	"A mouse (plural: mice) is a small rodent characteristically having a pointed snout, small rounded ears, a body-length scaly tail and a high breeding rate. The best known mouse species is the common house mouse ( <i>Mus musculus</i> ). It is also a popular pet. In some places, certain kinds of field mice are locally common. They are known to invade homes for food and shelter. Primarily nocturnal animals, mice compensate for their poor eyesight with a keen sense of hearing, and rely especially on their sense of smell to locate food and avoid predators."@en	<a href="http://dbpedia.org/resource/Mouse">http://dbpedia.org/resource/Mouse</a>
3	"Mouse Retinal Whole Mounts and Quantification of Vasculature Protocol"	"Mouse"	"A mouse (plural: mice) is a small rodent characteristically having a pointed snout, small rounded ears, a body-length scaly tail and a high breeding rate. The best known mouse species is the common house mouse ( <i>Mus musculus</i> ). It is also a popular pet. In some places, certain kinds of field mice are locally common. They are known to invade homes for food and shelter. Primarily nocturnal animals, mice compensate for their poor eyesight with a keen sense of hearing, and rely especially on their sense of smell to locate food and avoid predators."@en	<a href="http://dbpedia.org/resource/Mouse">http://dbpedia.org/resource/Mouse</a>

## Retrieve reagent names, their manufacturers and homepages

```
PREFIX sp: <http://purl.org/net/SMARTprotocol#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX obi: <http://purl.obolibrary.org/obo/OBI_>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX CHEBI: <http://purl.obolibrary.org/obo/CHEBI_>
```

```
SELECT ?name (group_concat(?manufacturerName;
separator=" , ") as ?manufacturers) (group_concat(?homepage; separator=" , ") as ?whereTobuy)
WHERE {
    ?reagent a CHEBI:33893 .
    ?reagent sp:hasName ?nameUri .
    ?nameUri rdf:value ?name .
    ?reagent obi:0000304 ?manufacturer .
    ?manufacturer sp:hasName ?manufacturerNameUri .
    ?manufacturerNameUri rdf:value ?manufacturerName .
    ?reagent foaf:homepage ?homepage .
} GROUP BY ?name
```

<http://bit.do/googledocs-sparql-sp>

<https://smartprotocols.github.io/queries/>

	name	manufacturers	whereTobuy
1	"Goat serum"	"Sigma-Aldrich"	"http://www.sigmaldrich.com/catalog/product/sigma/g9023"
2	"Sodium azide"	"Sigma-Aldrich"	"http://www.sigmaldrich.com/catalog/product/sial/s2002"
3	"Triton X-100"	"Sigma-Aldrich"	"http://www.sigmaldrich.com/catalog/product/sial/x100"
4	"Paraformaldehyde (PFA)"	"Electron Microscopy Sciences"	"https://www.fishersci.com/shop/products/16-paraformaldehyde-aq-solutn/50980487"
5	"Human MDA-MB-231 cell line"	"ATCC"	"https://www.lgcstandards-atcc.org/Products/All/HTB-26.aspx?geo_country=gb"

- Linked data model was agreed
- We have a working model for workflows
- We have generated linked data
- We are answering queries that are meaningful for the biologist –user stories
- We have identified how to systematically enrich our gazetteers
  - Web scraping and domain expert annotation

- Scraping many sources , e.g. journals, manufacturers web catalogs, databases
  - Grow our gazetteers
- Produce a lot of linked data representations for our corpus of experimental protocols

# Specification of experimental protocols



**Schema.org** describes “types” of information, which then have “properties”.

- Example: “recipe” is a type that has properties like “cookTime”, “cookingMethod”, “recipeYield”.

**Bioschemas** is working on specifications to improve the description of generic types in life sciences.

## **Goal:**

Extending schema.org to include the metadata proposed for documenting experimental protocols.

## **Deliverables:**

Registry of protocols using schema.org via bioschemas.



Olga Ximena Giraldo  
[ogiraldo@fi.upm.es](mailto:ogiraldo@fi.upm.es)  
[@olgaxgiraldo](https://twitter.com/olgaxgiraldo)

# Identification of linkable elements

**[Abstract]** The protocol describes the procedure of total RNA isolation from cells of the cyanobacterium *Synechocystis* sp. PCC 6803. This protocol is also applicable to *Synechococcus elongatus* PCC 6301, *Thermosynechococcus vulcanus*, and other unicellular and filamentous species of cyanobacteria that do not have thick polysaccharide-containing outer layers. For the latter, containing protocols should be adapted. The yield of RNA depends on optical density of cyanobacterial culture and may reach up to 10-20 µg of total RNA per 1 ml of cell culture. RNA isolated by this protocol can be used for Northern blot hybridization, RT-qPCR, microarrays and Next Generation Sequencing.

<http://bit.do/ncbiTaxon-PCC-6803>

NCBI Entrez PubMed Nucleotide Protein  
Search for:  AS complete name  lock   
Display: 3 levels using filter: none  
**Synechocystis sp. PCC 6803**  
Taxonomy ID: 1148  
Inherited blast name: cyanobacteria  
Rank: species  
Genetic code: Translation table 11 (Bacterial, Archaeal and Plant Plastid)  
Other names:  
synonym: *Synechocystis* sp. ATCC 27184  
synonym: *Synechocystis* sp. (ATCC 27184)  
synonym: *Aphanocapsa* sp. N-1  
synonym: *Aphanocapsa* sp. (strain N-1)  
includes: *Synechocystis* sp. PCC 6803 B  
includes: *Synechocystis* sp. PCC 6803 A  
equivalent name: *Synechocystis* sp. (strain PCC 6803)  
equivalent name: *Synechocystis* sp. (PCC 6803)

## Materials and Reagents

1. Guanidinium thiocyanate (Thermo Fisher Scientific, catalog number: BP221) (crossed out)
2. NaCitrate
3. N-Lauroylsarcosine (Sigma-Aldrich, catalog number: L9150)
4. β-mercaptoethanol (β-ME)
5. Glacial HAc
6. NaOH

<http://bit.do/pubchem-CID-65046>

PubChem OPEN CHEMISTRY DATABASE  
Compound Summary for CID 65046  
**Guanidinium Thiocyanate**  
STRUCTURE VENDORS PHARMACOLOGY LITERATURE BIOACTIVITIES  
PubChem CID: 65046  
Chemical Names: Guanidinium thiocyanate; Guanidinium rhodanide; Guanidine isothiocyanate; compd. with guanidine (1:1); Guanidine, monothiocyanate [More...](#)  
Molecular Formula: C<sub>2</sub>H<sub>6</sub>N<sub>4</sub>S  
Molecular Weight: 118.158 g/mol  
InChI Key: ZJYHGJYQJLLN-UHFFFAOYSA-N  
Safety Summary: Laboratory Chemical Safety Summary (LCSS)

Add new terms to our ontology

Add new synonyms to ontology terms

Terminology not curated

## Equipment

1. Vortex Mixer
2. Beckman high speed centrifuge with JS13.1 rotor or the like
3. 15-ml Corning centrifuge tube

# Goals for this meeting

- Extending schema.org to include the metadata proposed for documenting experimental protocols.

Metadata	Classification	Equivalent ‘property’ or ‘type’ in schema.org
<b>Metadata about the sample (input of the protocol)</b>		
sample	Type	-
<b>Metadata about the materials (or participants in a protocol)</b>		
equipment or supplies	Type	-
reagent	Type	-
software	Type	SoftwareApplication ( <a href="https://schema.org/SoftwareApplication">https://schema.org/SoftwareApplication</a> )
solution recipe	Type	-

Reuse metadata proposed in this meeting about samples

propose a new “type”, for representing lab devices

Reuse metadata proposed for this ‘type’

Metadata	Classification	Equivalent ‘property’ or ‘type’ in schema.org
<b>Metadata about the content</b>		
Advantage	property	-
limitation	property	-
application	property	-
purpose	property	-
Protocol instruction	property	-
outcome	property	we could reuse the extension term outcome ( <a href="https://health-lifesci.schema.org/outcome">https://health-lifesci.schema.org/outcome</a> )
provenance	property	isBasedOn ( <a href="https://schema.org/isBasedOn">https://schema.org/isBasedOn</a> )
troubleshooting	property	-
<b>Metadata about alert message</b>		
caution	property	-
Critical step	property	-
hint	property	-
Pause point	property	-
Storage conditions	property	storageRequirements ( <a href="https://schema.org/storageRequirements">https://schema.org/storageRequirements</a> )
Timing	property	totalTime ( <a href="https://schema.org/totalTime">https://schema.org/totalTime</a> ) Duration ( <a href="https://schema.org/Duration">https://schema.org/Duration</a> )