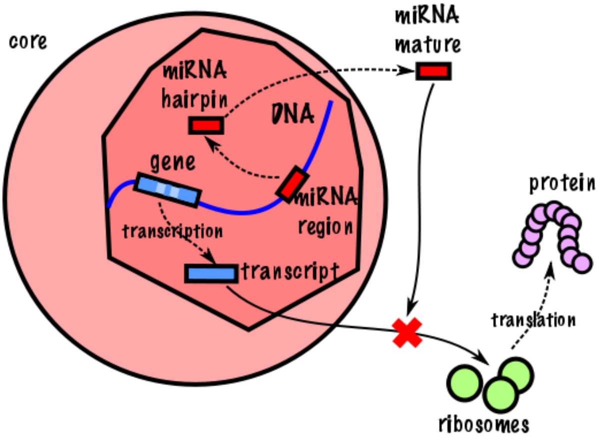
Publishing Diachronic Life Science Linked Data

Katerina Gkirtzou[1] Thanasis Vergoulis[1] Artemis Hatzigeorgiou[3]
Timos Sellis[2] Theodore Dalamagas[1]

1 ΙΠΣΥ IMIS 2 RMIT UNIVERSITY 3 UNIVERSITY OF THESSALY

SWAT4LS - 10th December, 2014

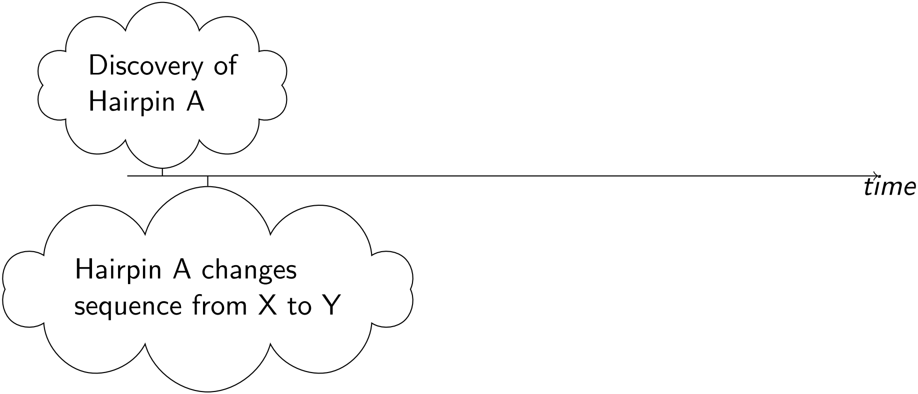
MicroRNA Biogenesis Path



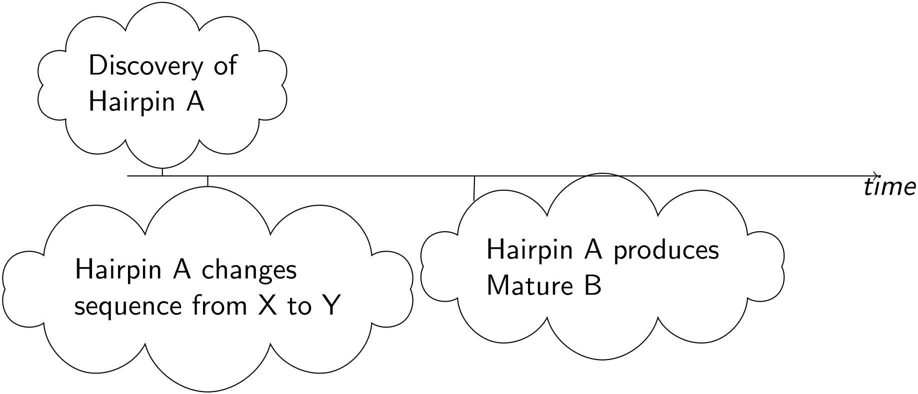
The problem

The problem

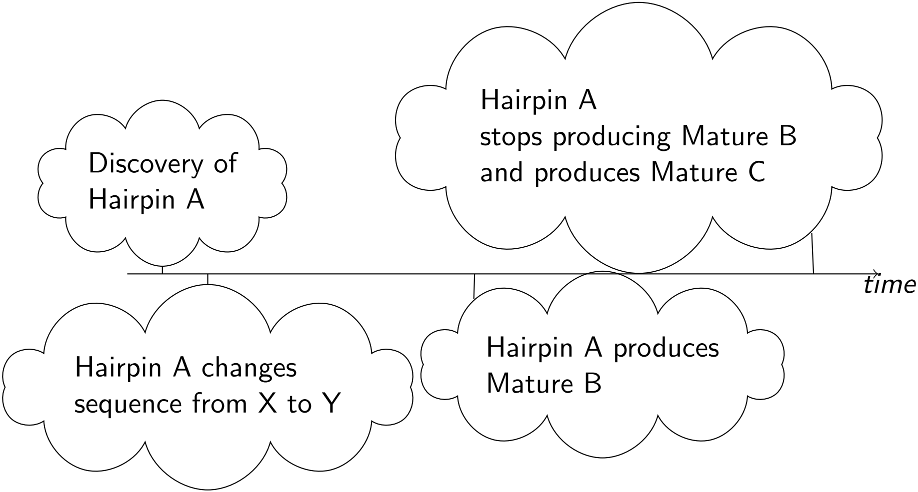
The problem


Discovery of Hairpin A
Hairpin A changes sequence from X to Y
time

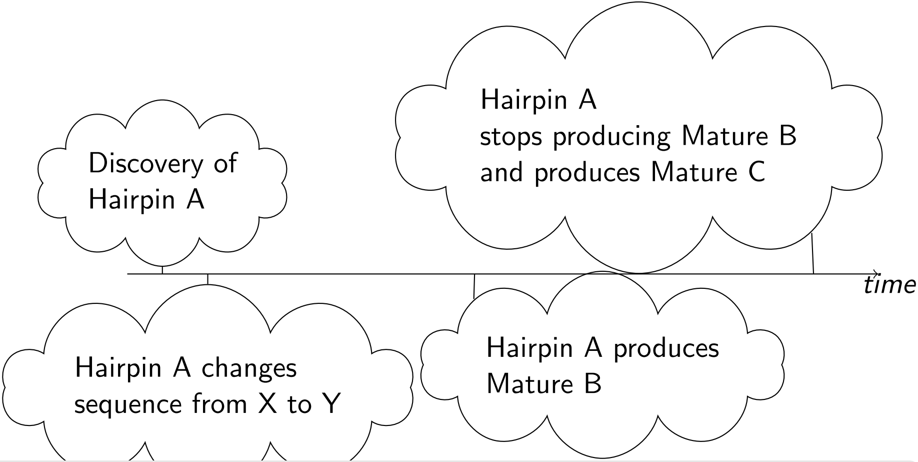
The problem


Discovery of Hairpin A
time
Hairpin A changes sequence from X to Y
Hairpin A produces Mature B

The problem


Discovery of Hairpin A
Hairpin A changes sequence from X to Y
Hairpin A produces Mature B
Hairpin A stops producing Mature B and produces Mature C
time

The problem



Requirements

1. Having access not only to up-to-date LD, but to any of the previous versions.
2. Being able to track the changes among versions, as well as their causes and effects.

MicroRNA Curated Data

DIANA tools - http://diana.imis.athena-innovation.gr/DianaTools/index.php

- miRBase Database - http://www.mirbase.org/
- Ensemble database - http://www.ensembl.org/index.html
- KEGG pathways database - http://www.genome.jp/kegg/pathway.html
- Manually curated bibliography related to miRNAs from Pubmed

MicroRNA Curated Data

DIANA tools - http://diana.imis.athena-innovation.gr/DianaTools/index.php

- miRBase Database - http://www.mirbase.org/
- Ensemble database - http://www.ensembl.org/index.html
- KEGG pathways database - http://www.genome.jp/kegg/pathway.html
- Manually curated bibliography related to miRNAs from Pubmed

Changes observed on miRBase database

Change Type	Description	Hairpin	Mature
NEW	A new miRNA is created	✓	✓
NAME	The miRNA changes its name	✓	✓
SEQUENCE	The miRNA changes its sequence	✓	✓
NAME/SEQUENCE	The miRNA changes both its name and sequence at the same time	✓	✓
DELETE	The miRNA is deleted	✓	✓
FORWARD	The miRNA change its miRBase accession id to a new one	✓	
ADD HAIRPIN PARENT	A new hairpin is added to the list of hairpins that produce a mature.	✓	✓
REMOVE HAIRPIN PARENT	A hairpin is removed from the list of hairpins that produce a mature.	✓	✓

Overiew of proposed change model based on RDF

1st Requirement

Having access not only to up-to-date scientific LD, but to any of the previous versions.

2nd Requirement

Being able to track the changes among versions, as well as their causes and effects.

Overiew of proposed change model based on RDF

1st Requirement

Having access not only to up-to-date scientific LD, but to any of the previous versions.

Proposed solution

Use versioned RDF entities and "version" properties

2nd Requirement

Being able to track the changes among versions, as well as their causes and effects.

Overiew of proposed change model based on RDF

1st Requirement

Having access not only to up-to-date scientific LD, but to any of the previous versions.

Proposed solution

Use versioned RDF entities and “version” properties

2nd Requirement

Being able to track the changes among versions, as well as their causes and effects.

Proposed Solution

Model changes as RDF properties and RDF entities.

Modeling versions

URI definition

1. General URI :
   `http://.../resource/class-name/resource-id`
2. Versioned URI :
   `http://.../resource/class-name/resource-id/timestamp`

Modeling versions

URI definition

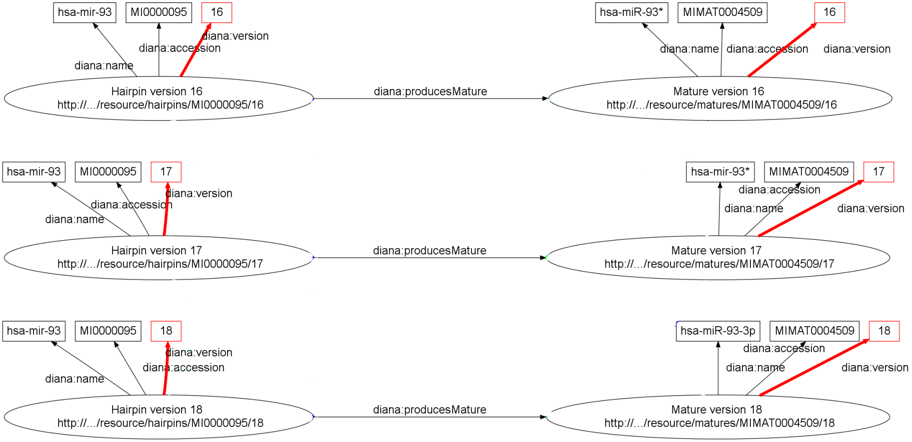
1. General URI : `http://.../resource/class-name/resource-id`
2. Versioned URI : `http://.../resource/class-name/resource-id/timestamp`

Version Properties

- `:label`, with the constant value `"now"`,
- `:version`, with version timestamp values,
- `:previousVersion`, and
- `:nextVersion`.

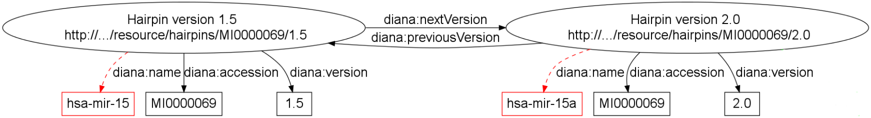
Versioning example

Versioning example

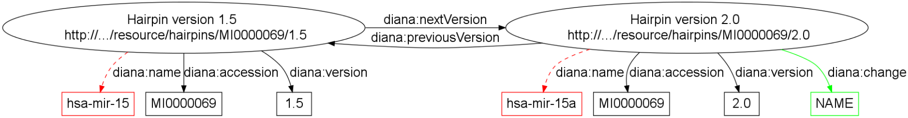


Versioning example

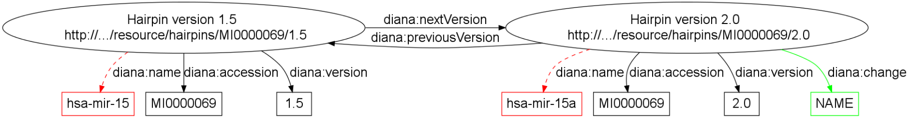
Change on Property to Literal



Change on Property to Literal

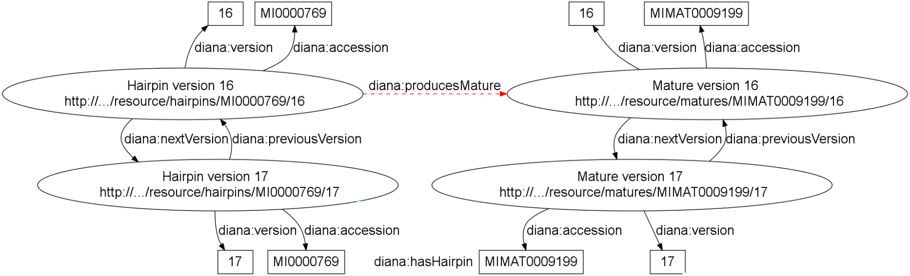


Change on Property to Literal

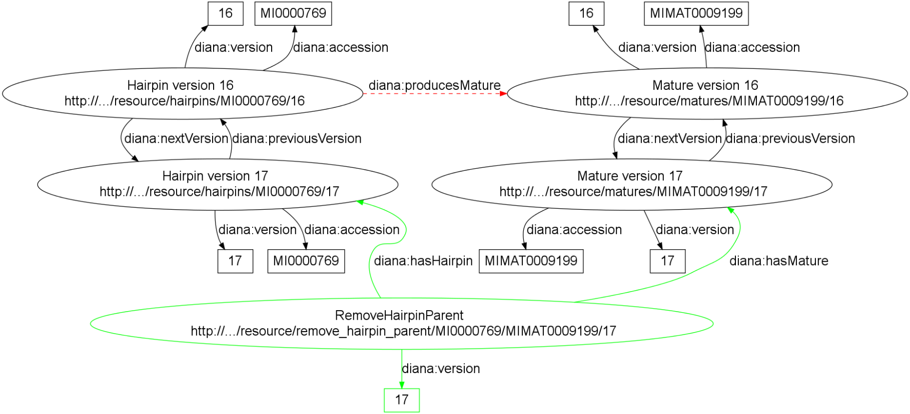


Change Type	**Description**	**Hairpin**	**Mature**
NEW	A new miRNA is created	✓	✓
NAME	The miRNA changes its name	✓	✓
SEQUENCE	The miRNA changes its sequence	✓	✓
NAME/SEQUENCE	The miRNA changes both its name and sequence at the same time	✓	✓
DELETE	The miRNA is deleted	✓	✓
FORWARD	The miRNA change its miRBase accession id to a new one	✓	

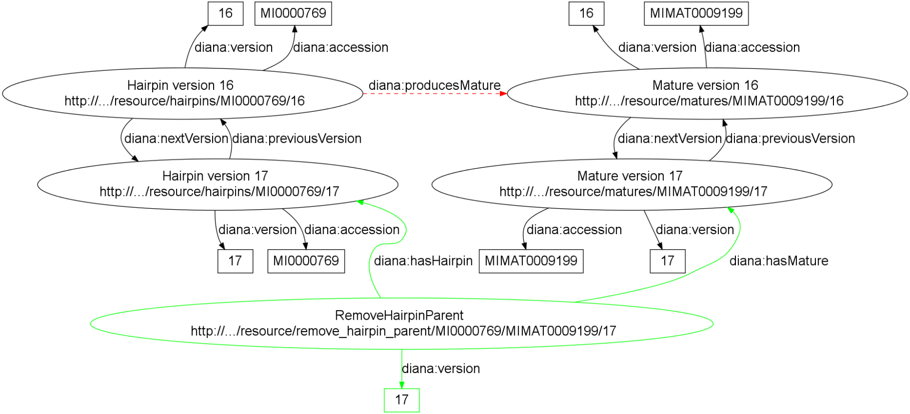
Change on Property to Resource



Change on Property to Resource



Change on Property to Resource



Change Type	Description
ADD HAIRPIN PARENT	A new hairpin is added to the list of hairpins that produce a mature.
REMOVE HAIRPIN PARENT	A hairpin is removed from the list of hairpins that produce a mature.

Query on current data

Query

Retrieve Hairpins, their produced Matures and their Targets as predicted by microT-ANN program from the current version of miRNA LOD.

SPARQL

```
SELECT ?hairpin ?mature ?target WHERE
{
  ?hairpin rdf:type diana:Hairpin.
  ?hairpin diana:label "now".
  ?hairpin diana:producesMature ?mature.
  ?interaction rdf:type diana:Interaction.
  ?interaction diana:hasMature ?mature.
  ?interaction diana:hasTarget ?target.
  ?interaction diana:application "microT-ANN (v4.0)".
}
```

Query on out-of-date data

Query

Retrieve Hairpins, their produced Matures and their Targets as predicted by microT-ANN program from version “13.0” of miRBAse of miRNA LOD.

SPARQL

```
SELECT ?hairpin ?mature ?target WHERE
{
  ?hairpin rdf:type diana:Hairpin.
  ?hairpin diana:version "13.0".
  ?hairpin diana:producesMature ?mature.
  ?interaction rdf:type diana:Interaction.
  ?interaction diana:hasMature ?mature.
  ?interaction diana:hasTarget ?target.
  ?interaction diana:application "microT-ANN (v4.0)".
}
```

Query of a Change on Property to Literal

Query

Retrieve all Hairpins that change their name over their lifetime. Also get the old and the new value of the name as well as the version the change occured.

SPARQL

```
SELECT ?hairpin ?version ?old_name ?new_name WHERE
{
  ?hairpin rdf:type diana:Hairpin.
  ?hairpin diana:change "NAME".
  ?hairpin diana:version ?version.
  ?hairpin diana:name ?new_name.
  ?hairpin diana:previousVersion ?hairpin_prev.
  ?hairpin_prev diana:name ?old_name.
}
```

Query of a Change on Property to Resource

Query

Retrieve all pairs of Hairpins and Matures that were related via the relationship that a hairpin produces a mature and the relationship ended until version “18” of miRBase.

SPARQL

```
SELECT DISTINCT ?hairpin ?version ?mature WHERE
{
  ?connection rdf:type diana:removeHairpinParent.
  ?connection diana:version ?version.
  FILTER (?version <= "18").
  ?connection diana:hasHairpin ?hairpin.
  ?connection diana:hasMature ?mature.
}
```

Coclusions - Future Work

Summing up

1. We proposed a model that captures change and version information based on RDF that allows
   - Exploration of entities with up-to-date information.
   - Exploration of out-of-date resource descriptions of entities and navigation between previous and next versions of the resources.
   - Modeling changes as first citizen RDF resources.
2. Curated data of the miRNA world from well-known miRNA databases are exported as LOD using the proposed model.
3. All services are available at `http://diana.imis.athena-innovation.gr/lod`.

Future work

- Interlink our miRNA LOD with LD from EBI RDF Platform.
- Provide keyword search querying capabilities on our framework.

Related Work

- This work is an extension of [Dalamagas 12]
  1. Use of a different database schema
  2. Integration of new miRNA-related datasets.
- The Memento framework [de Sompel 10] handles versions of LD by attaching time-specific attributes.
- [Umbrich 10a] focuses on monitoring and examining the dynamics of LD.
- [Popitsch 10] and [Volz 09], deal with the problem of broken links.
- [Umbrich 10b] studies the aspects of discovery, change detection at several granularity level andd escription of the changes.
- Evolutionary Terminology Auditing (ETA) [Ceu 10] examines how terminology changes reflect evolutions in the underlying domain.
- Named graphs [Carroll 05] allow easier provenance tracking and trustness warranties.

Acknowledgments

Acknowledgments

Work has been supported by LODGOV project, Research Programme ARISTEIA (EXCELLENCE), General Secretariat for Research and Technology, Ministry of Education, Greece and the European Social Fund.

European Union
European Social Fund

OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION & RELIGIOUS AFFAIRS, CULTURE & SPORTS
MANAGING AUTHORITY

NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND

Co-financed by Greece and the European Union

- Jeremy J. Carroll, Christian Bizer, Pat Hayes & Patrick Stickler.
  Named graphs.
  Web Semantics: Science, Services and Agents on the World Wide Web, vol. 3, no. 4, pages 247 – 267, 2005.

- *Applying Evolutionary Terminology Auditing to SNOMED CT*.
  In AMIA 2010 Proceedings, pages 96–100, 2010.

- Theodore Dalamagas, Nikos Bikakis, George Papastefanatos, Yannis Stavrakas & Artemis G. Hatzigeorgiou.
  Publishing life science data as linked open data: the case study of miRBase.
  In WOD, pages 70–77, 2012.

- Herbert Van de Sompel, Robert Sanderson, Michael L. Nelson, Lyudmila Balakireva, Harihar Shankar & Scott Ainsworth.
  An HTTP-Based Versioning Mechanism for Linked Data.
  CoRR, 2010.

Niko Popitsch & Bernhard Haslhofer.
DSNotify: handling broken links in the web of data.
In WWW, pages 761–770, 2010.

Jrgen Umbrich, Michael Hausenblas, Aidan Hogan, Axel Polleres & Stefan Decker.
Towards Dataset Dynamics: Change Frequency of Linked Open Data Sources.
In LDOW, 2010.

Jrgen Umbrich, Boris Villazn-Terrazas & Michael Hausenblas.
Dataset Dynamics Compendium: A Comparative Study.
In COLD, volume 665, 2010.

Julius Volz, Christian Bizer, Martin Gaedke & Georgi Kobilarov.
Discovering and Maintaining Links on the Web of Data.
In ISWC 2009, pages 650–665, 2009.