



HAL
open science

Using annotations for R function management, MOQA : Méta-données et Ontologies pour la Qualité des Annotations

Anne Tireau, Caroline Domerg, Olilvier Corby, Juliette Fabre, Catherine Faron Zucker, Emilie Gennari, Alexandre Granier, Isabelle Mirbel, Vincent Negre, Pascal Neveu

► To cite this version:

Anne Tireau, Caroline Domerg, Olilvier Corby, Juliette Fabre, Catherine Faron Zucker, et al.. Using annotations for R function management, MOQA : Méta-données et Ontologies pour la Qualité des Annotations. JOBIM: Journées Ouvertes en Biologie, Informatique et Mathématiques, Sep 2010, Montpellier, France. 36 p. hal-02824208

HAL Id: hal-02824208

<https://hal.inrae.fr/hal-02824208v1>

Submitted on 6 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



for Using annotations function management

A. Tireau
C. Domerg
O. Corby
J. Fabre
C. Faron-Zucker
É. Gennari
A. Granier
I. Mirbel
V. Nègre
P. Neveu

6th September 2010

MOQA 2010

Context

Approach

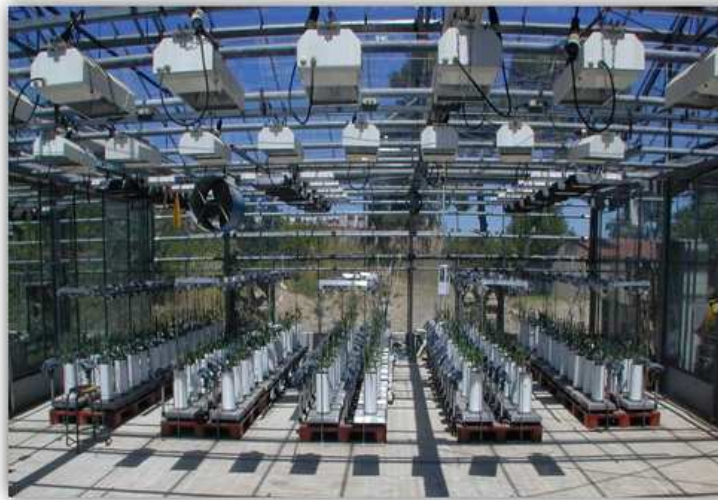
Ontology Overview

Demonstration

Conclusion

INRA - LEPSE:

Laboratory of Plant Ecophysiology under Environmental Stresses



- Plant adaptation to climatic change
- Controlled environment
- High-throughput phenotyping
- Information Systems (Phenopsis, Phenodyn, Cincalli)

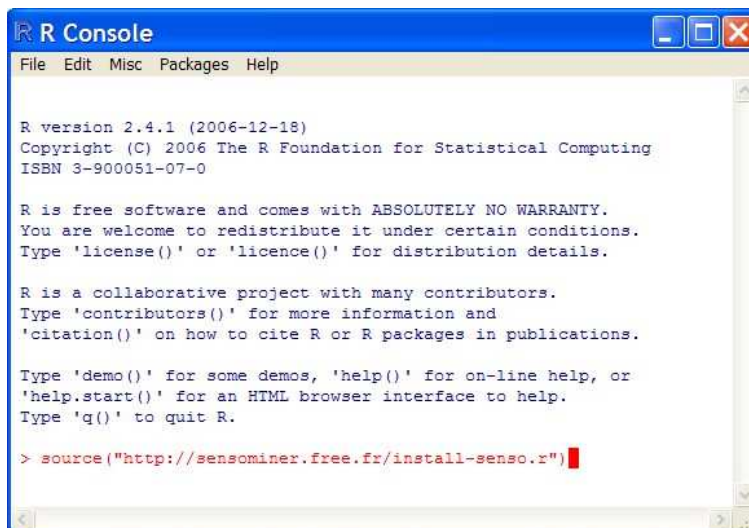


tool developments for data management with



What is ?

- Computer programming language and environment for statistical computing and graphics
- Used by a large researcher community
- Open source (GNU project) and multi- platforms (Linux, Windows, MacOS)
- Easy to extend with user-written functions.



```
R Console
File Edit Misc Packages Help

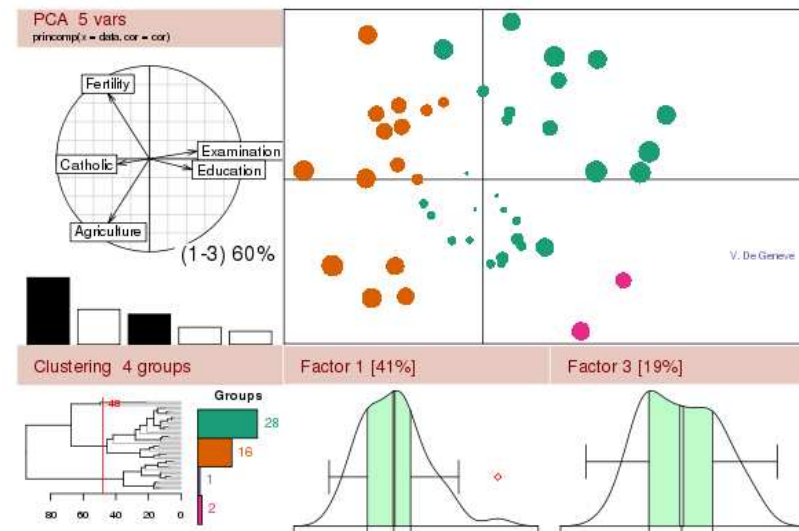
R version 2.4.1 (2006-12-18)
Copyright (C) 2006 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> source("http://sensominer.free.fr/install-senso.r")
```



Using for agronomy research

● Information Systems:

- ✗ Online monitoring and data visualisation
- ✗ Data checking, validation and insertion (online/offline data)

```
library(RODBC)

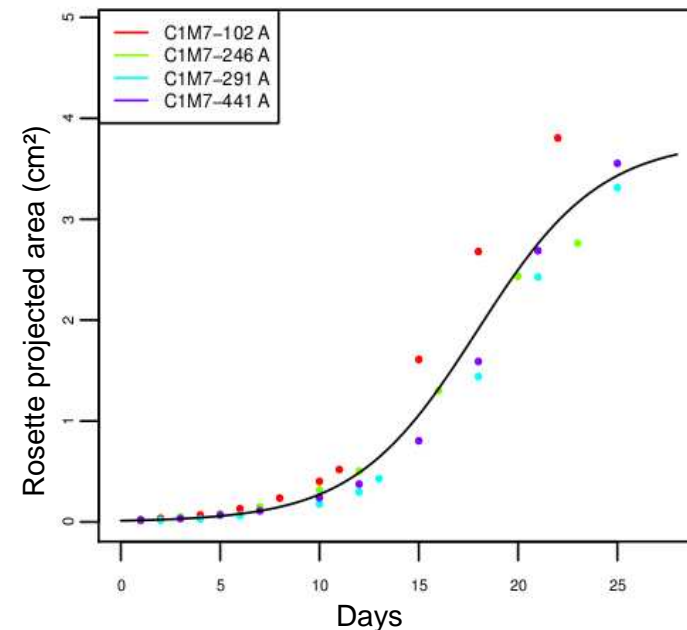
channel <- odbcConnect(dsn="phenopsis",
uid="user", pwd="pwd", case="nochange")

req_sql <- paste("LOAD DATA LOCAL
INFILE '", dataTmp, "' INTO TABLE
MesureMeteo(idChambre, date,
idVariable, valeur);", sep="")

sqlQuery(channel, req_sql)
```

● Data Analysis:

- ✗ Elaborated data computation
- ✗ Statistical analysis
- ✗ Modelling



High production of
 functions

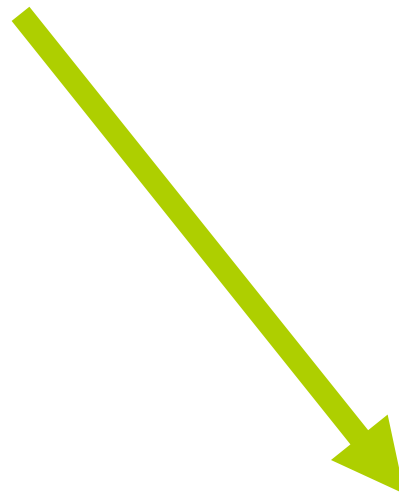


Many authors
and turnover

High production of
 **functions**



Many authors
and turnover



How to share, capitalize, organize and promote these functions?

AIMS

Store and organize the functions
Give an easy and long-term access

IDEAS

Create an ontology to describe R functions
Provide a new kind of repository with reasoning and powerful search tools

TOOLS

W3C Semantic Web Technologies

Ontology

Formal description of concepts and relations between concepts

Examples of concept: **Rfunction** **Argument** **Person**

Examples of relation: **hasArgument** **isANewVersionOf** **hasAuthor**

→ Provides a **controlled** vocabulary

→ Designed to be understood by **computers**

RDF, RDFS and OWL: W3C standards to write ontologies

RDF

Model, language and syntax to describe resources and their relations

RDFS

Structures RDF resources (class and property hierarchies)

OWL

Built on top of RDFS, allows to define property features (transitivity, symmetry, etc.)

RDF

 Resource Description Framework

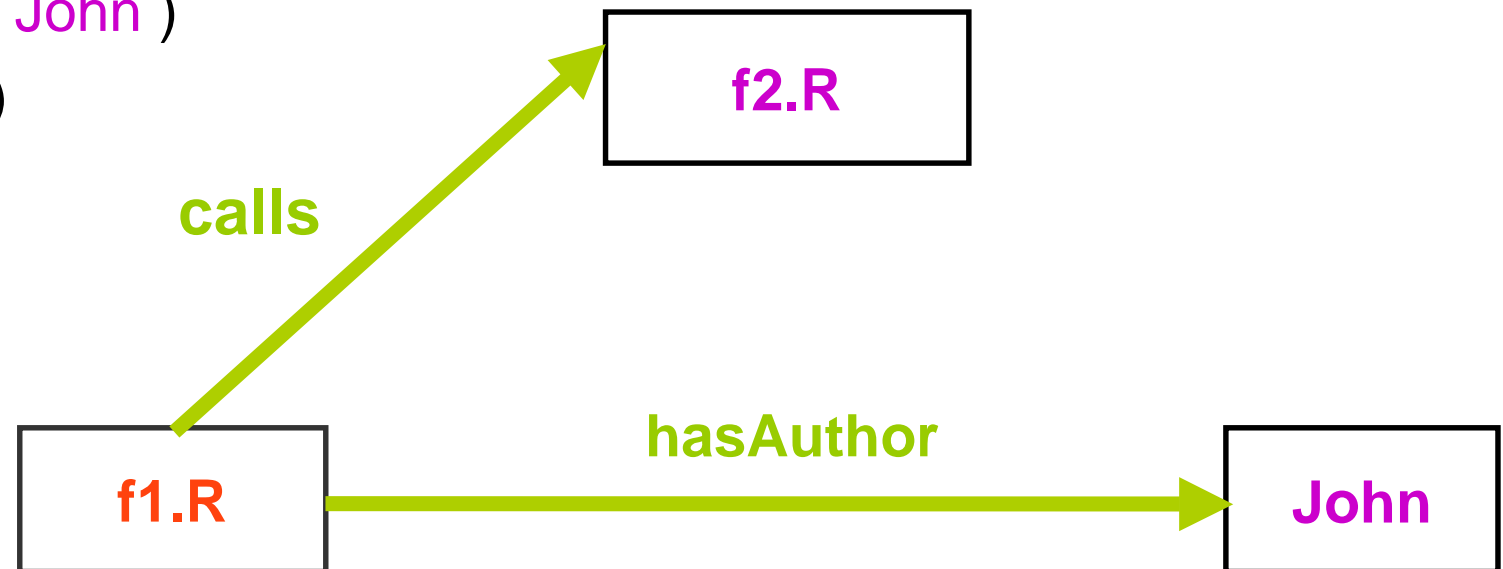
An example about R function description:

→ R function attributes and properties

→ Relations between R functions

(f1.R , hasAuthor , John)

(f1.R , calls , f2.R)



RDF is a semantic graph model

RDF, RDFS and OWL: W3C standards to write ontologies

RDF

Model, language and syntax to describe resources and their relations

RDFS

Structures RDF resources (class and property hierarchies)

OWL

Built on top of RDFS, allows to define property features (transitivity, symmetry, etc.)

RDF Schema

Provides elements to **structure** RDF resources such as:

→ Class hierarchy

→ Property restrictions (domain, range)

Example of class: domain and range definition



RDF, RDFS and OWL: W3C standards to write ontologies

RDF

Model, language and syntax to describe resources and their relations

RDFS

Structures RDF resources (class and property hierarchies)

OWL

Built on top of RDFS, allows to define property features (transitivity, symmetry, etc.)

OWL

Ontology Web Language

→ Built on top of RDFS

→ Allows to define : transitivity, symmetry, inverse of, etc. properties

➔ Provides powerful description of concepts and their relationships

Example of OWL: INVERSE OF



Ontology querying

RDF/OWL files

Ontology and annotations

```
<owl:ObjectProperty rdf:about="#couldBeUsedAfter">  
  <rdfs:range rdf:resource="#Rfunction"/>  
  <rdfs:domain rdf:resource="#Rfunction"/>  
  <owl:inverseOf rdf:resource="#couldBeUsedBefore"/>  
</owl:ObjectProperty>
```

SPARQL

Query

Language

for RDF/OWL

Ontology querying

RDF/OWL files: Ontology and annotations

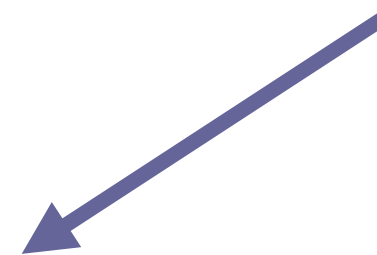
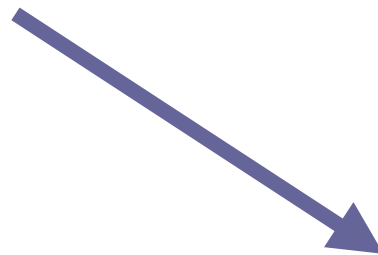
```
<owl:ObjectProperty rdf:about="#couldBeUsedAfter">  
  <rdfs:range rdf:resource="#Rfunction"/>  
  <rdfs:domain rdf:resource="#Rfunction"/>  
  <owl:inverseOf rdf:resource="#couldBeUsedBefore"/>  
</owl:ObjectProperty>
```

SPARQL

Query

Language

for RDF/OWL



CORESE (*INRIA – Edelweiss*)

Software and engine to :

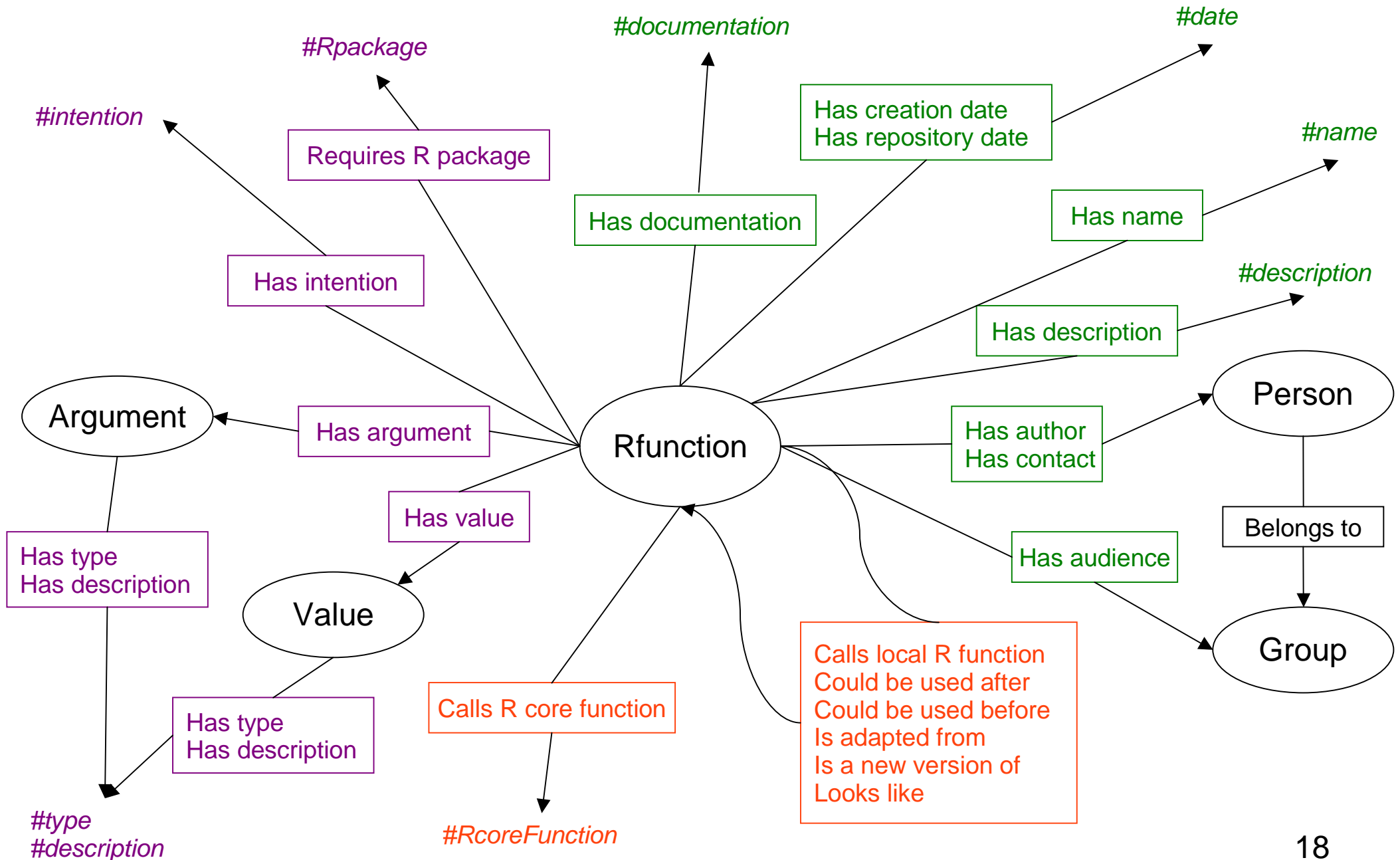
→ infer and to run rules

→ perform SPARQL queries

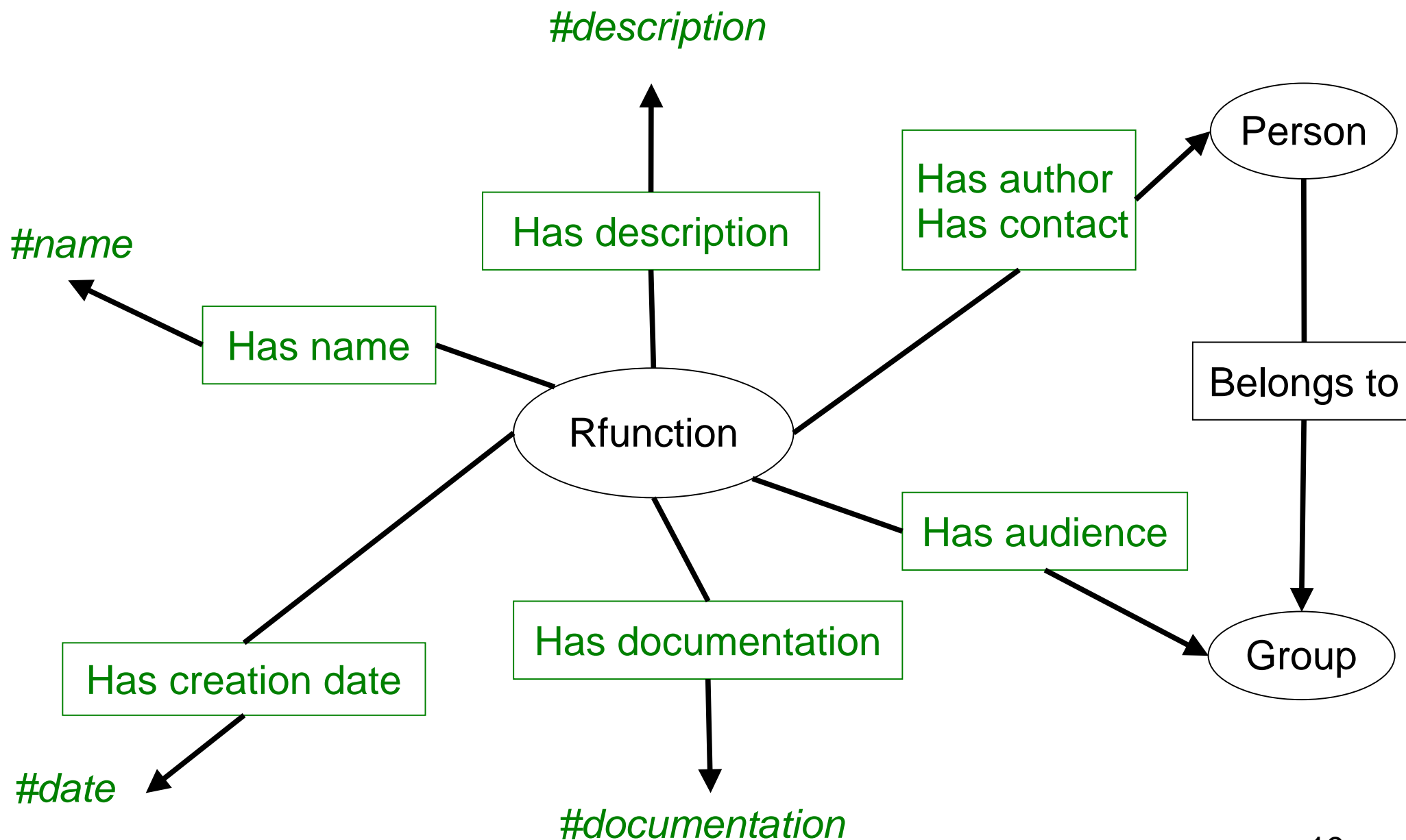


**Query
results**

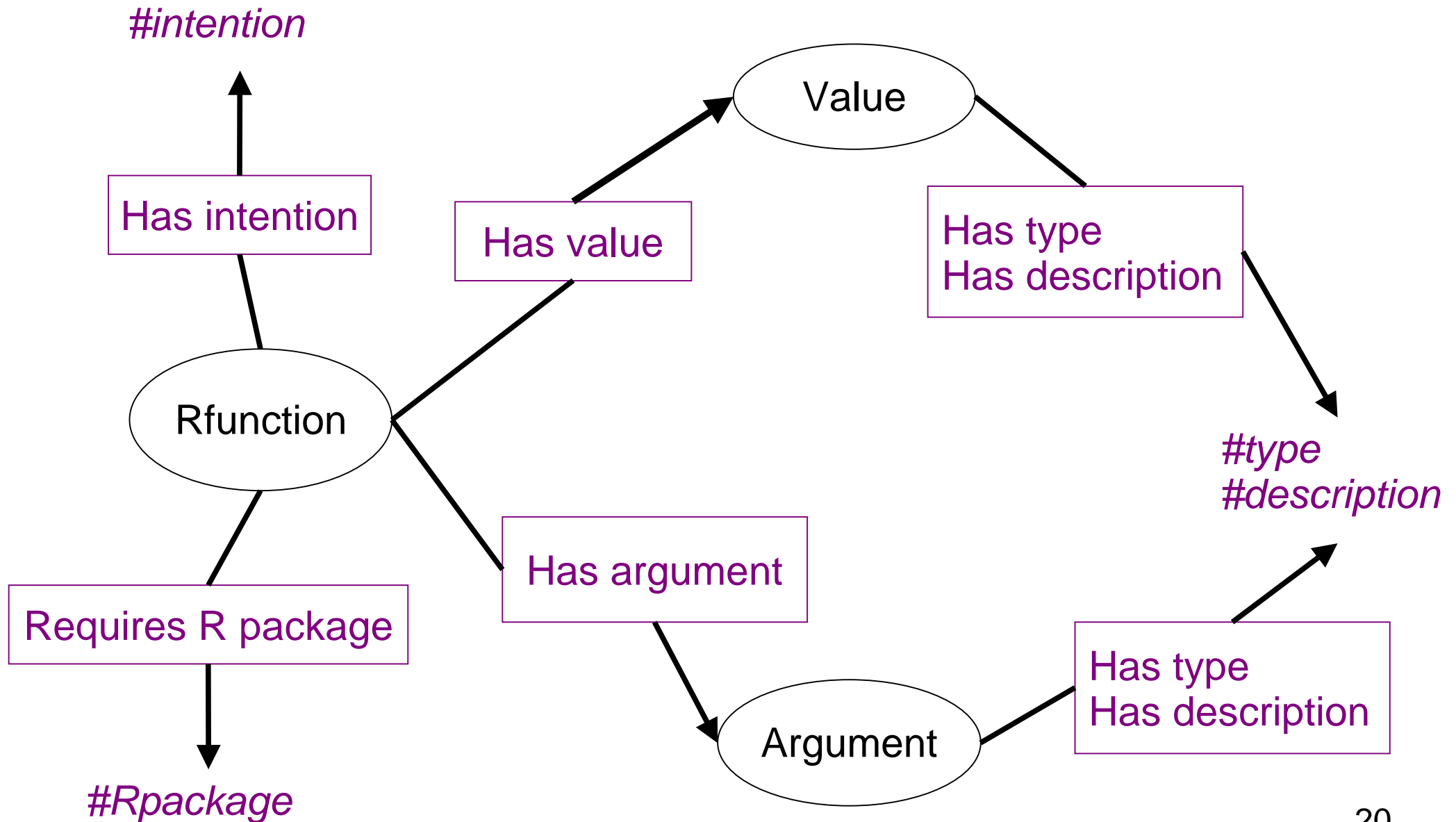
Global view



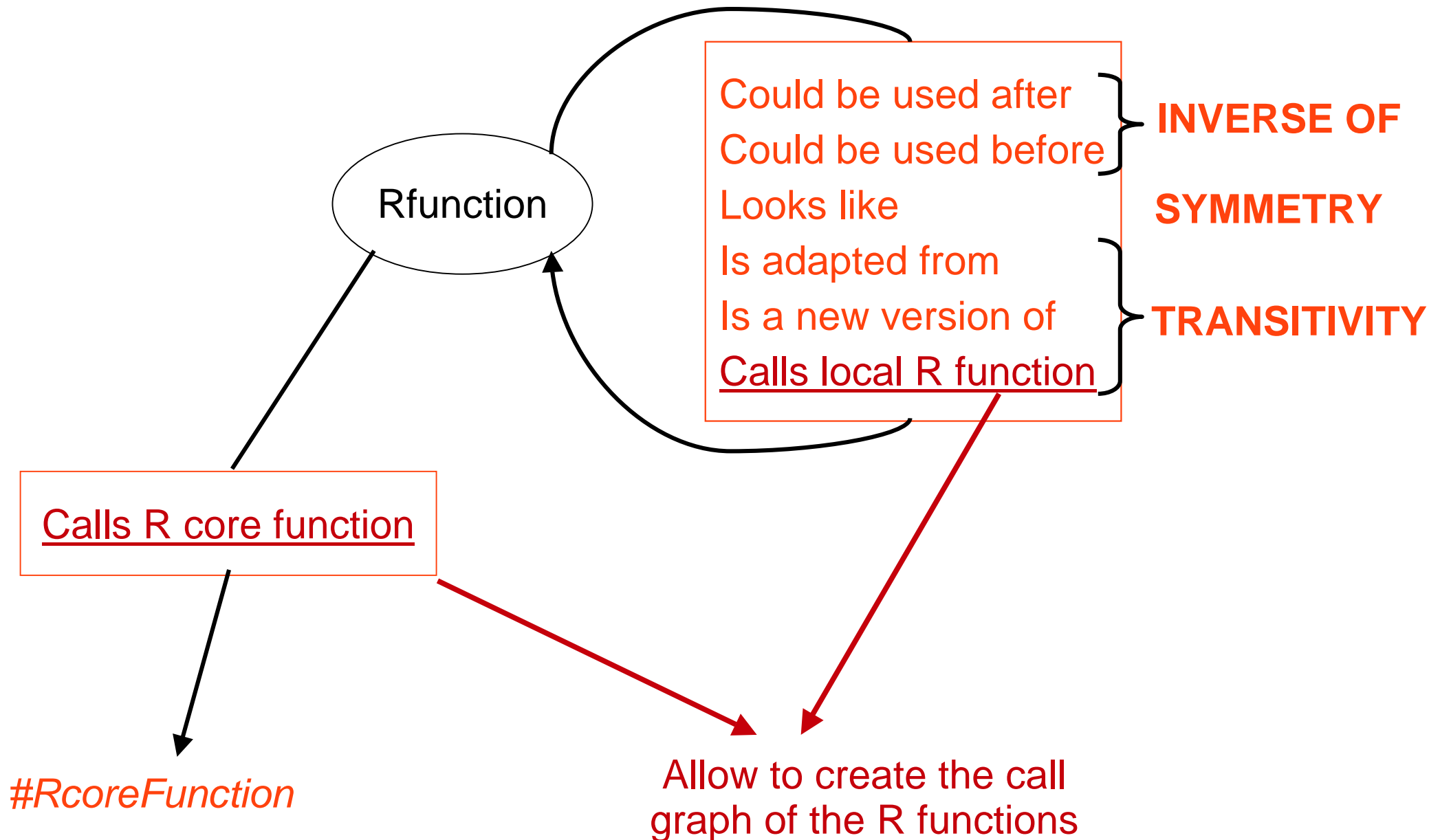
General description of R functions



Detailed description of R functions



Relations with other R functions



OWL source code

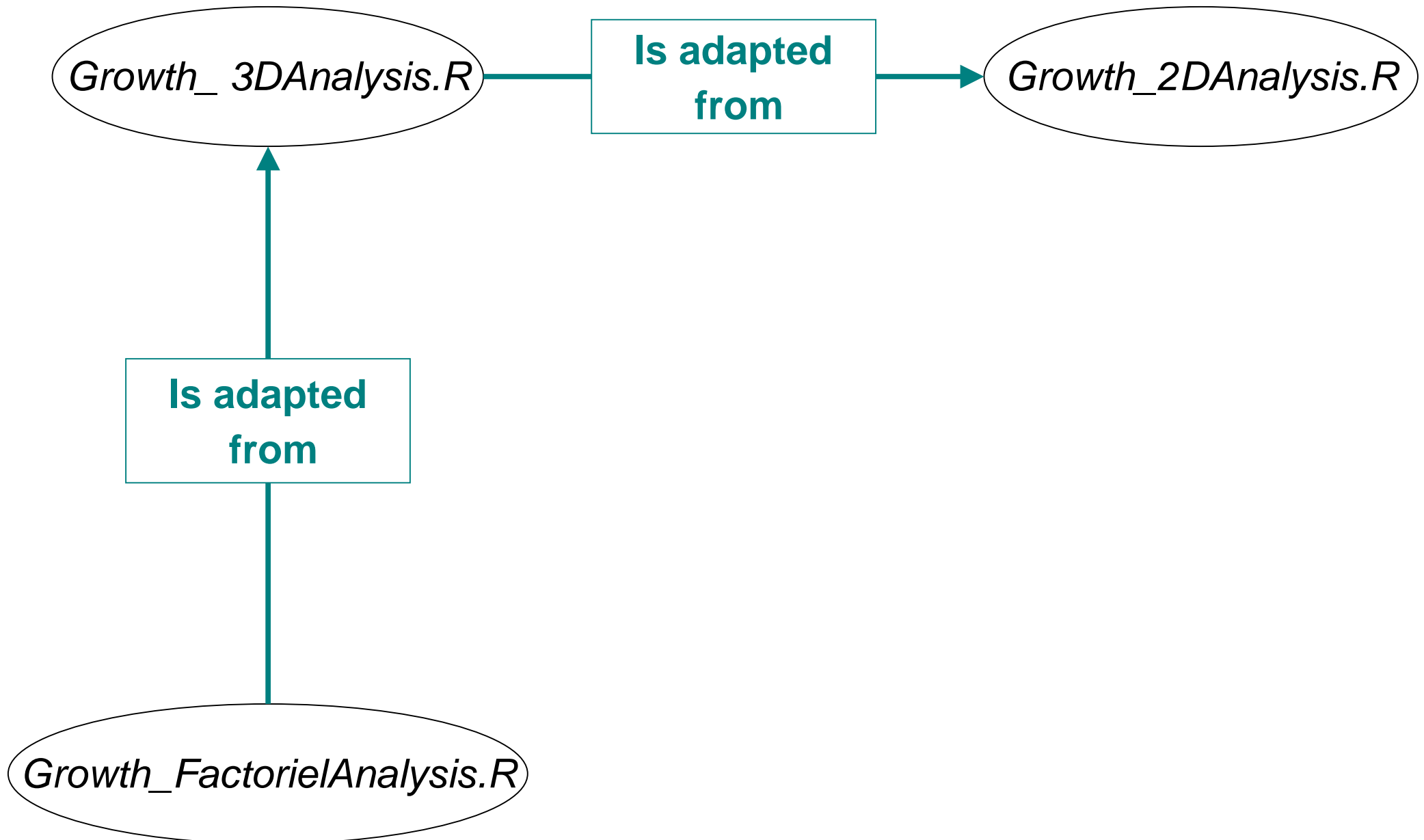
INVERSE OF

```
<owl:ObjectProperty rdf:about="#couldBeUsedAfter">
  <rdfs:comment rdf:datatype="&rdfs;Literal">
    If the R function could be used after another R function
  </rdfs:comment>
  <rdfs:range rdf:resource="#Rfunction"/>
  <rdfs:domain rdf:resource="#Rfunction"/>
  <owl:inverseOf rdf:resource="#couldBeUsedBefore"/>
</owl:ObjectProperty>
```

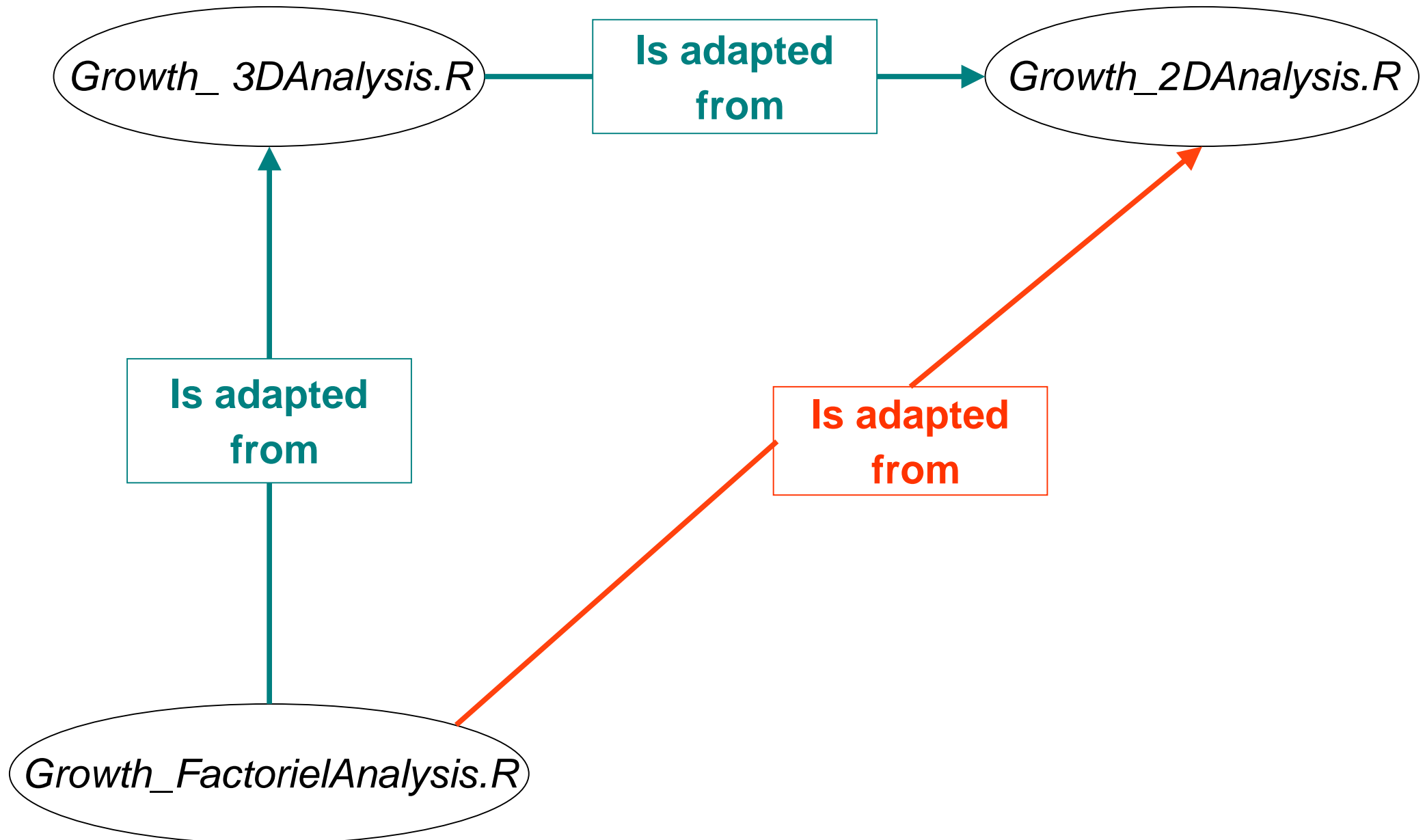
TRANSITIVITY

```
<owl:ObjectProperty rdf:about="#hasLocalCall">
  <rdf:type rdf:resource="&owl;TransitiveProperty" />
  <rdfs:comment rdf:datatype="&rdfs;Literal">
    Local Call function</rdfs:comment>
  <rdfs:range rdf:resource="#Rfunction"/>
  <rdfs:domain rdf:resource="#Rfunction"/>
</owl:ObjectProperty>
```

Relations deduced by Inference : **TRANSITIVITY**



Relations deduced by Inference : **TRANSITIVITY**



Relations deduced by Inference : **INVERSE OF**



LER_Computation.R



LER_Computation.R

Relations deduced by Inference : **INVERSE OF**



LER_Computation.R



LER_Plot.R



LER_Computation.R

**Could be
used after**

LER_Plot.R

`<couldBeUsedAfter rdf:resource="#LER_Computation.R />`

Relations deduced by Inference : **INVERSE OF**



LER_Computation.R



LER_Plot.R



LER_Computation.R

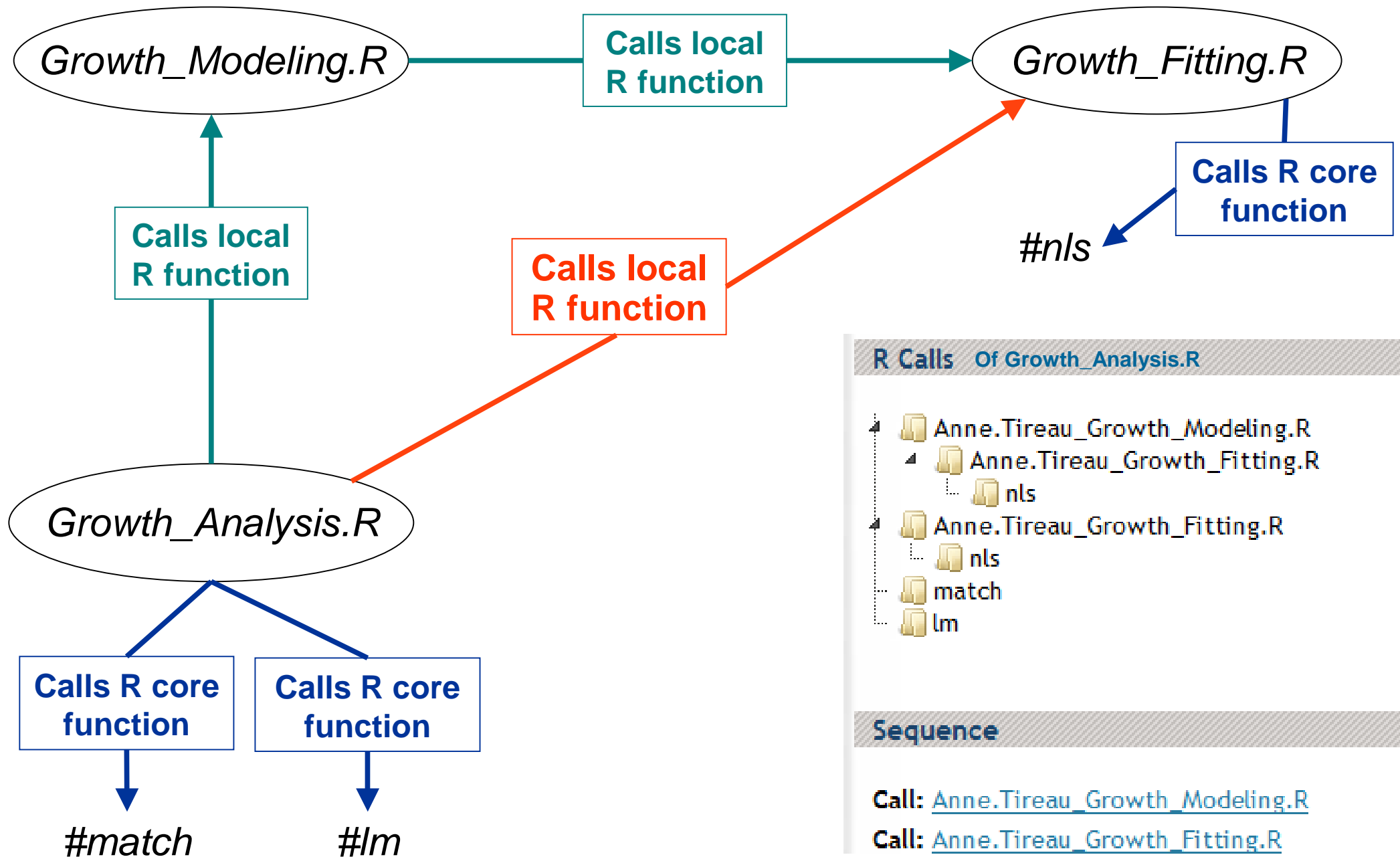
**Could be
used after**

**Could be
used before**

LER_Plot.R

`<couldBeUsedAfter rdf:resource="#LER_Computation.R />`

Relations with other R functions: *Call Graph*



R Calls Of Growth_Analysis.R

- └ Anne.Tireau_Growth_Modeling.R
 - └ Anne.Tireau_Growth_Fitting.R
 - └ nls
- └ Anne.Tireau_Growth_Fitting.R
 - └ nls
- └ match
- └ lm

Sequence

Call: [Anne.Tireau_Growth_Modeling.R](#)

Call: [Anne.Tireau_Growth_Fitting.R](#)

**Web Interface
Demonstration**

Conclusions

- Users find this repository relevant (efficient search, easy annotating)
- Semantic Web tools allow reasoning for an 'intelligent' repository
- Model and application are easy to adapt to:
 - ↳ other fields of research,
 - ↳ other programming languages,
 - ↳ mathematical models.

Conclusions

- Users find this repository relevant (efficient search, easy annotating)
- Semantic Web tools allow reasoning for an 'intelligent' repository
- Model and application are easy to adapt to:
 - ↳ other fields of research,
 - ↳ other programming languages,
 - ↳ mathematical models.

Prospects

- Add relations between functions and reports, articles, etc.
- Add concept of R package
- Improve automatic extraction from R function documentation

Thank you
Questions ?

GUI for edition and creation of annotations

- A few minutes thanks to pre-filled forms
- Generation and storage of OWL file

1 - General

*Fields followed by * should be filled!*

Name of the function * :

The name should be of the following form: *MyFunction.R*

Description * :

Multiple selection or unselection: use <Ctrl>

Author(s):	Person(s) to contact :
<input type="text" value="Anne.Pellegrino"/> <input type="text" value="Benoit.Boussuge"/> <input type="text" value="Bertrand.Muller"/> <input type="text" value="Caroline.Domerg"/> <input type="text" value="Christian.Fournier"/> <input type="text" value="Christine.Granier"/>	<input type="text" value="Christian.Fournier"/> <input type="text" value="Christine.Granier"/> <input type="text" value="Eric.Lebon"/> <input type="text" value="Bertrand.Muller"/> <input type="text" value="Vincent.Negre"/> <input type="text" value="Anne.Pellegrino"/>

Creation date (ex: 2010-11-26) * :

3 - Detailed description

All the following informations are optional:

Audience:	Information system(s) concerned:	Intention(s) of the function:
<input type="text" value="Statistician"/> <input type="text" value="Ecophysiologicalist"/> <input type="text" value="Genetician"/> <input type="text" value="Operator"/>	<input type="text" value="Cincalli"/> <input type="text" value="Phenodyn"/> <input type="text" value="Phenopsis"/>	<input type="text" value="DataChecking"/> <input type="text" value="DataTransformation"/> <input type="text" value="DatabaseConnection"/> <input type="text" value="Modeling"/> <input type="text" value="StatisticalAnalysis"/> <input type="text" value="Visualisation"/>

Execution time of the function (short, medium, long):

R package(s) required separated by ';' (eg: ade4; lattice):

Main R functions called in the function separated by ';' (eg: lm; curve; nls):

Argument description

Describe all the different arguments of the function:

Argument 1	Name: <input type="text" value="dataset"/>	Type: <input type="text" value="array"/>	Description: <input type="text"/>
-------------------	--	--	-----------------------------------

2 - Uploads

Upload the R script (.R):

Function consultation card

Vera.Georgescu_LERvalidation.R

 [Download R files](#)

Description

The general function of visualisation, automatic and manual correction of the Leaf Elongation Rate kinetics measured on the Phenodyn platform. This function runs on R version 2.6.2

Arguments

■ LERvalidation.R_graph

Type: logical

Description: boolean for graphic mode (for manual correction)

■ LERvalidation.R_finnuit

Type: scalar

Description: the hour of end of the night.

Audience

Ecophysiologicalist

Authors

Vera.Georgescu

Contacts

Vincent.Negre

Creation date

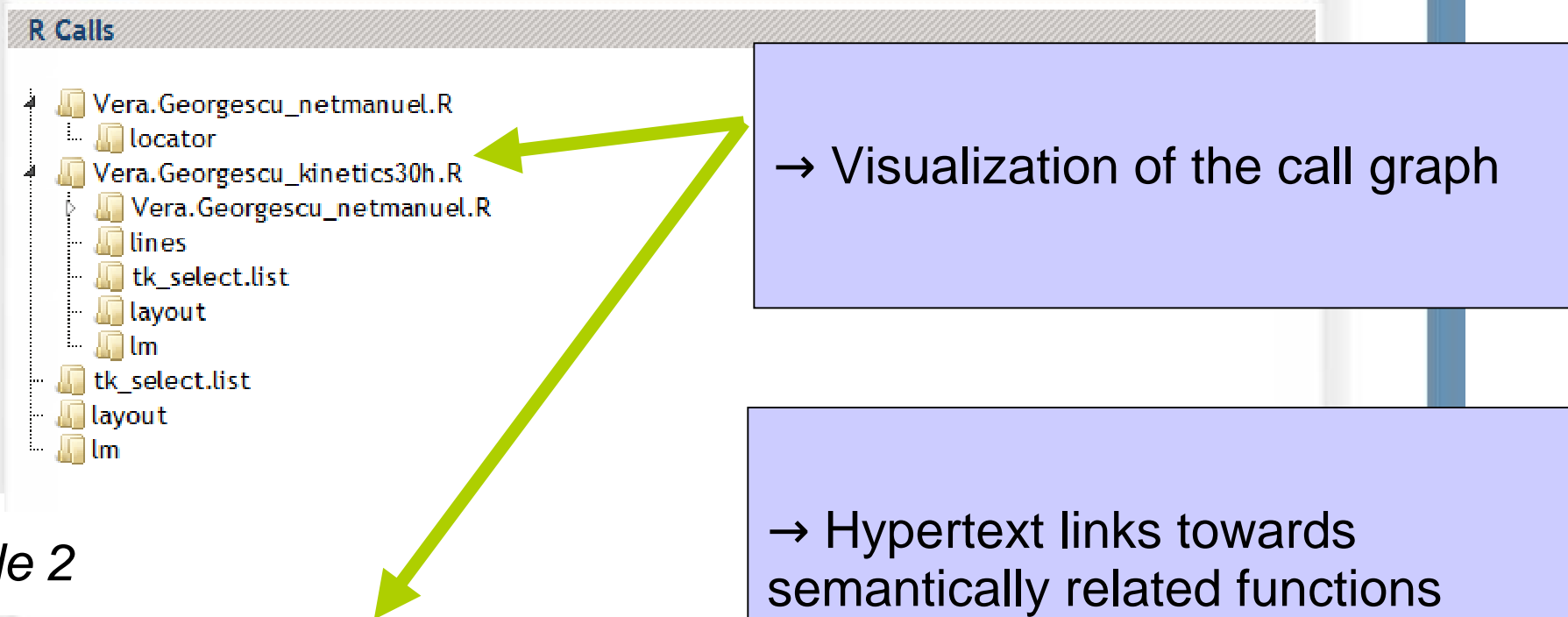
2008-02-01

→ Information about the function (author, arguments, intentions, etc.)

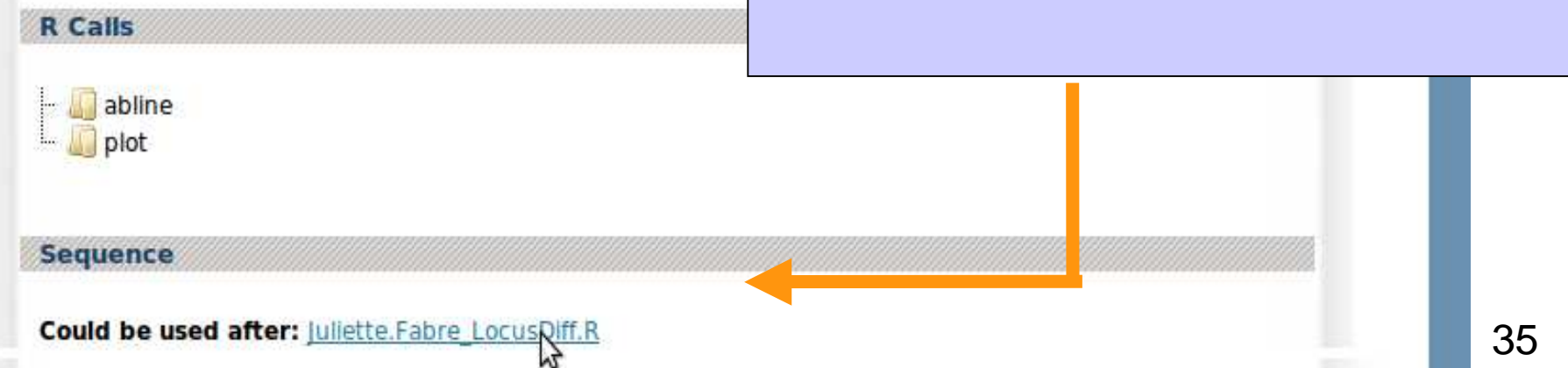
→ Download of R function and associated files (documentation, datasets, etc.)

Function consultation card

Example 1




Example 2




Powerful search tools

Build a *SPARQL* request adding conditions on the properties


Search an R-Function

Basic Search (if you know the name of the function)
 Advanced Search

```
PREFIX OntologyR: select ?fonction ?description where { ?fonction OntologyR:isDedicatedTo OntologyR:Phenodyn ?fonction OntologyR:hasIntention OntologyR:Visualisation ?fonction OntologyR:hasDescription ?description }
```

 There are 4 functions matching your request:

Name	Description
Vera.Georgescu_LERvalidation.R	The general function of visualisation, automatic and manual correction of Elongation Rate kinetics measured on the Phenodyn platform. This function runs on R version 2.6.2
Vera.Georgescu_kinetics30h.R	The function gives a representation of Phenodyn leaf elongation rate kinetics for one night and the following day and night (about 30 h). It performs and represents simple regressions on the nights. It allows manually the LER data and performs the new regressions when data have been invalidated. It displays a selection list that proposes to correct the data, see the following day or come back to the previous day. This function runs on R version 2.6.2

Example: search the functions

→ Dedicated to the Information System 'Phenodyn' and with an intention of Visualization

→ That could be used after 'ImportData.R'

→ That call the R core function 'anova'