# 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

# Using annotations for R 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

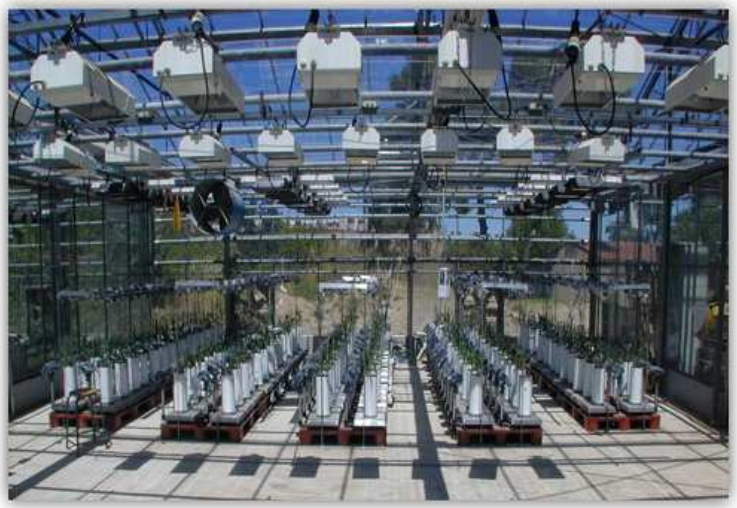
Ontology Overview

Demonstration

Conclusion

# INRA - LEPSE:

## Laboratory of Plant Ecophysiology under Environmental Stresses

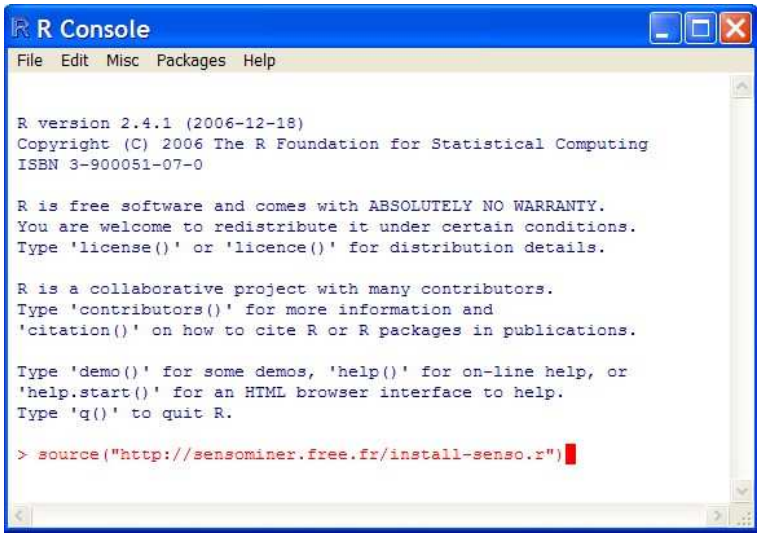



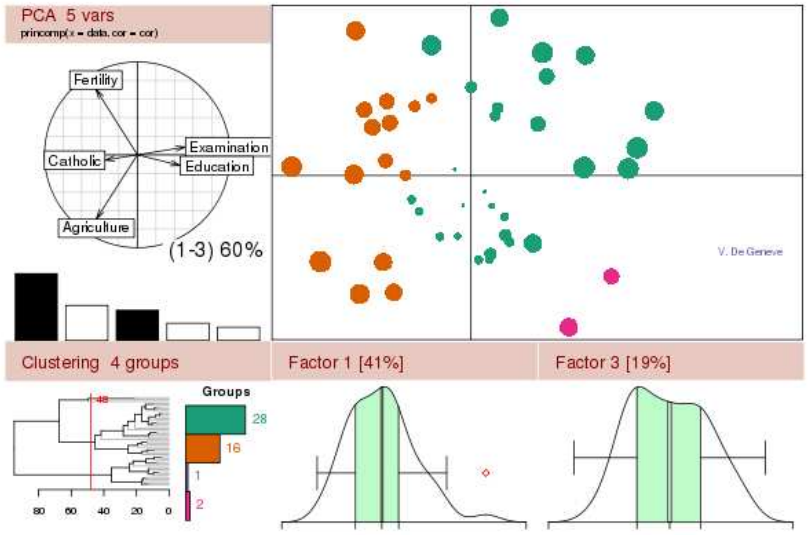

- Plant adaptation to climatic change
- Controlled environment
- High-throughput phenotyping
- Information Systems (Phenopsis, Phenodyn, Cincalli)

→ tool developments for data management with

# What is R ?

- 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.




http://cran.r-project.org/

# Using R for agronomy research

- **Information Systems**:
  - Online monitoring and data visualisation
  - Data checking, validation and insertion (online/offline data)

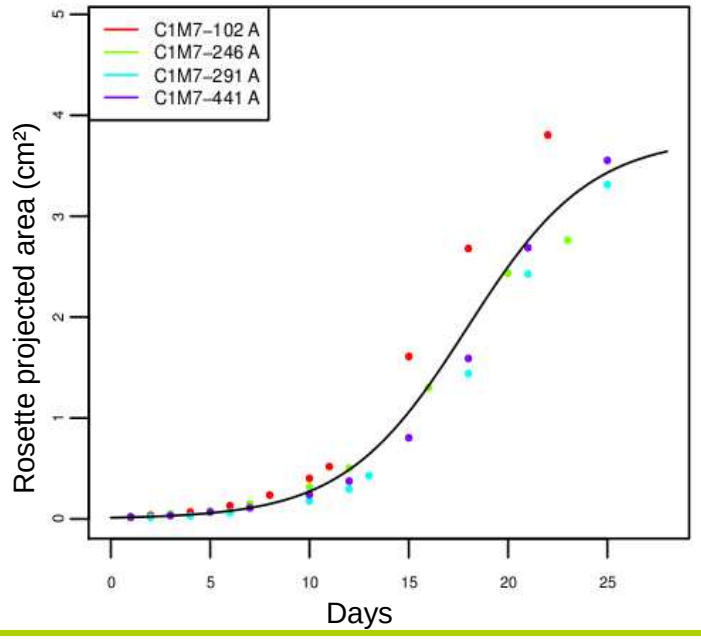
```
library(RODBC)

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** | **RDFS** | **OWL** |
| --- | --- | --- |
| Model, language and syntax to describe resources and their relations | Structures RDF resources (class and property hierarchies) | Built on top of RDFS, allows to define property features (transitivity, symmetry, etc.) |

# RDF Resource Description Framework

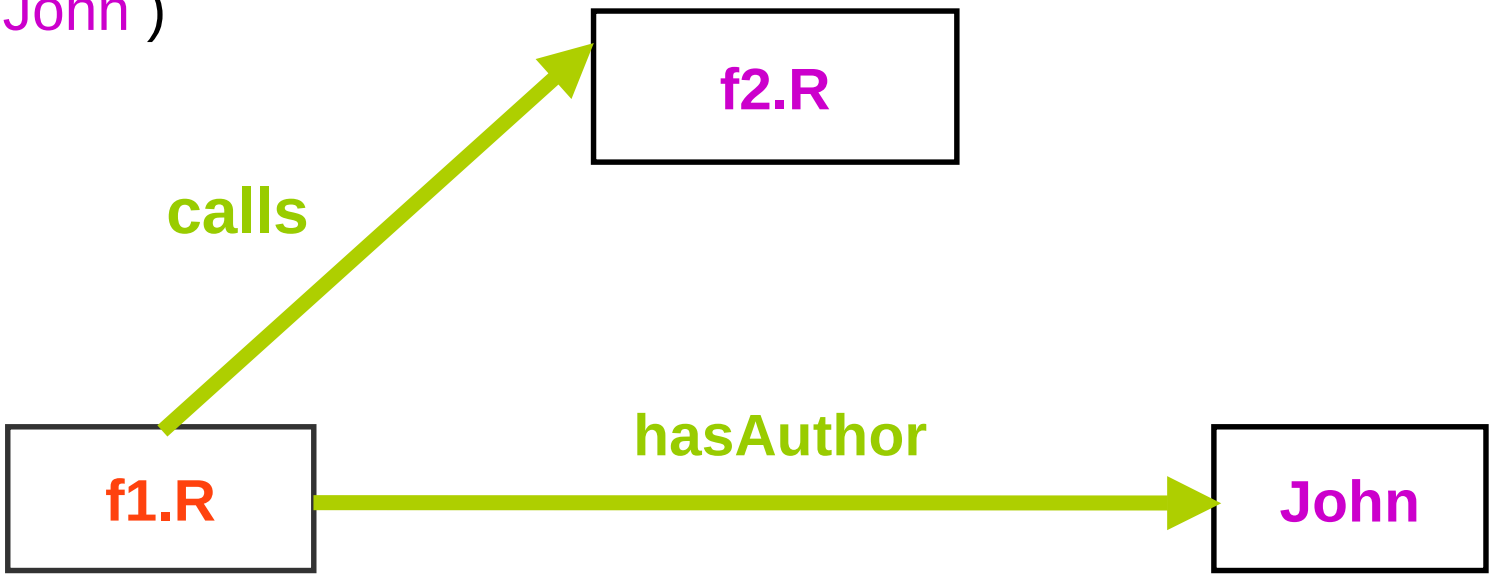
An example about R function description:

→ 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 | **RDFS** | OWL |
| --- | --- | --- |
| Model, language and syntax to describe resources and their relations | Structures RDF resources (class and property hierarchies) | 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 | RDFS | **OWL** |
| --- | --- | --- |
| Model, language and syntax to describe resources and their relations | Structures RDF resources (class and property hierarchies) | 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*)

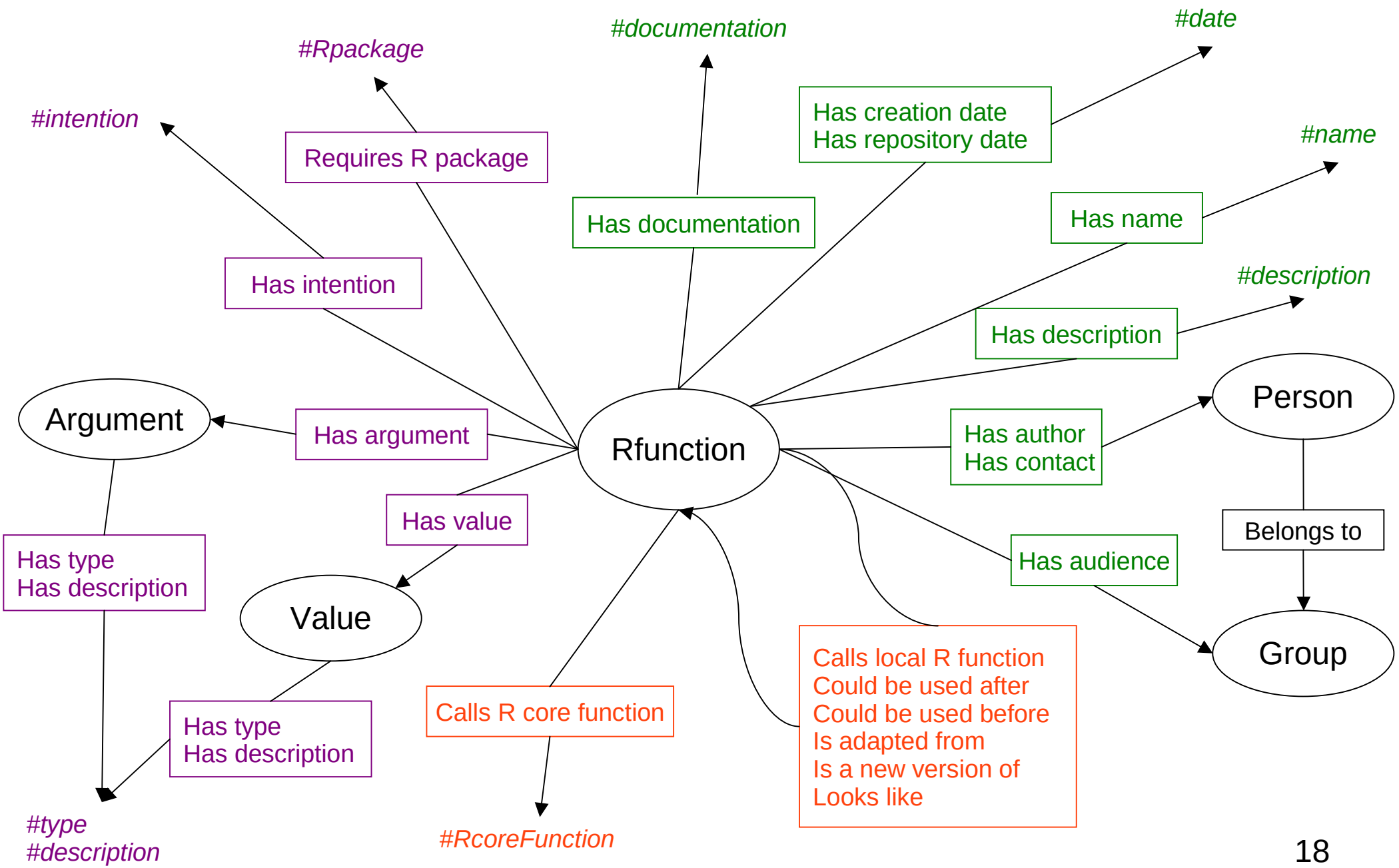
Software and engine to :

→ infer and to run rules

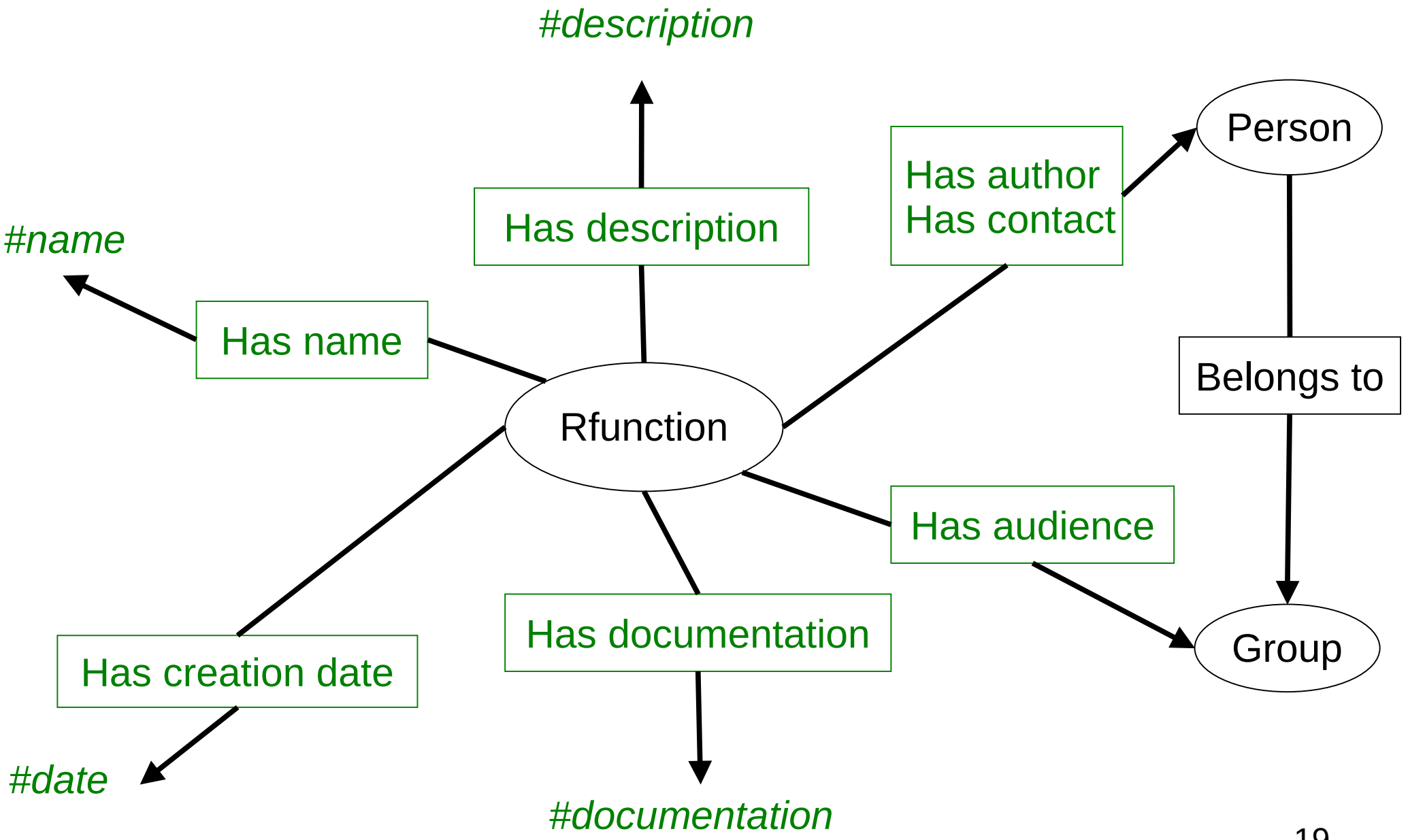
→ perform SPARQL queries

***Query results***

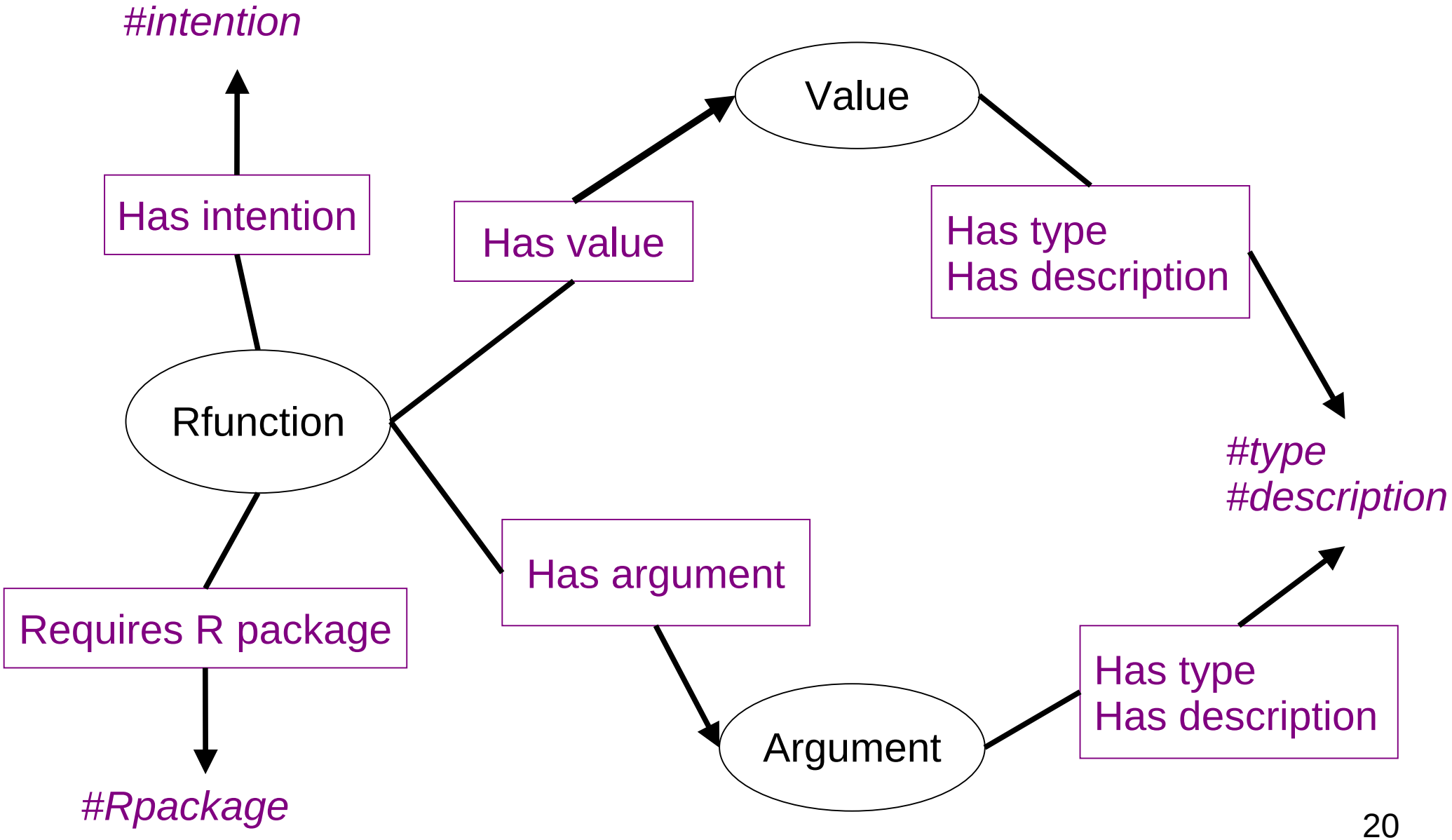
# Global view


#Rpackage
#documentation
#date
#intention
Requires R package
Has creation date
Has repository date
#name
Has documentation
Has name
Has intention
#description
Has description
Argument
Has argument
Rfunction
Has author
Has contact
Person
Has value
Has type
Has description
Value
Has audience
Belongs to
Group
Calls local R function
Could be used after
Could be used before
Is adapted from
Is a new version of
Looks like
Has type
Has description
Calls R core function
#type
#description
#RcoreFunction

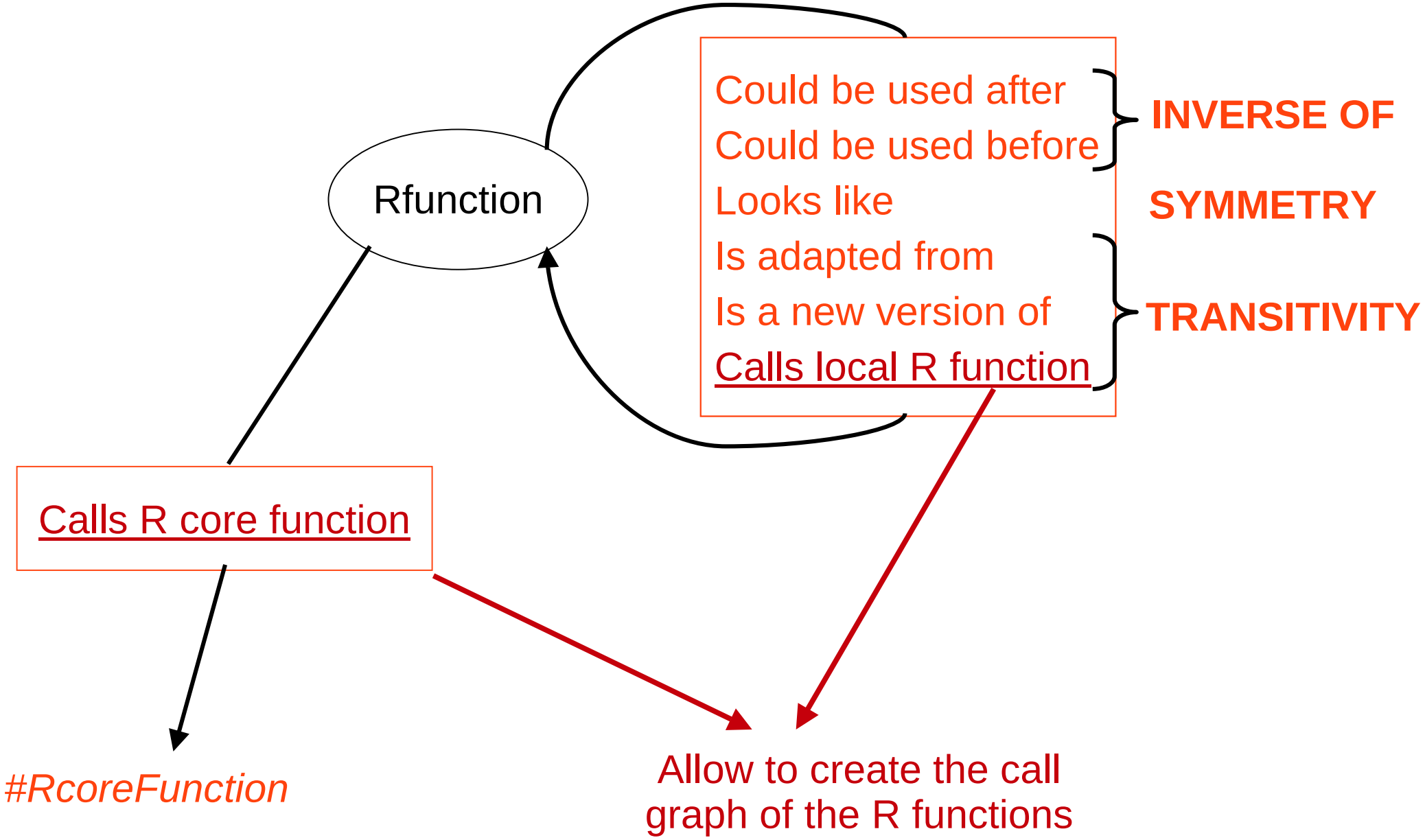
# General description of R functions



# Detailed description of R functions



# Relations with other R functions



# OWL source code

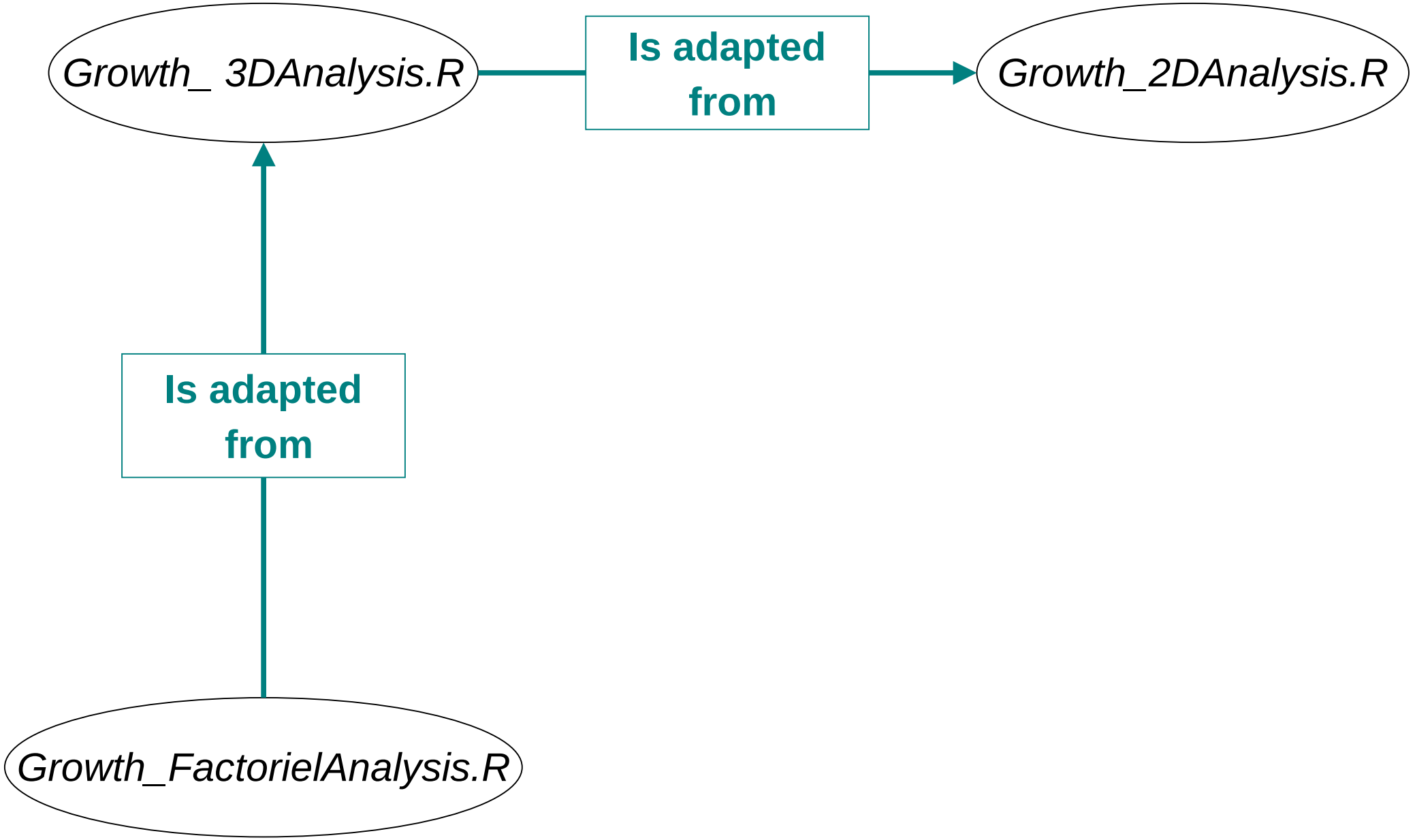
## INVERSE OF

```
<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>
```

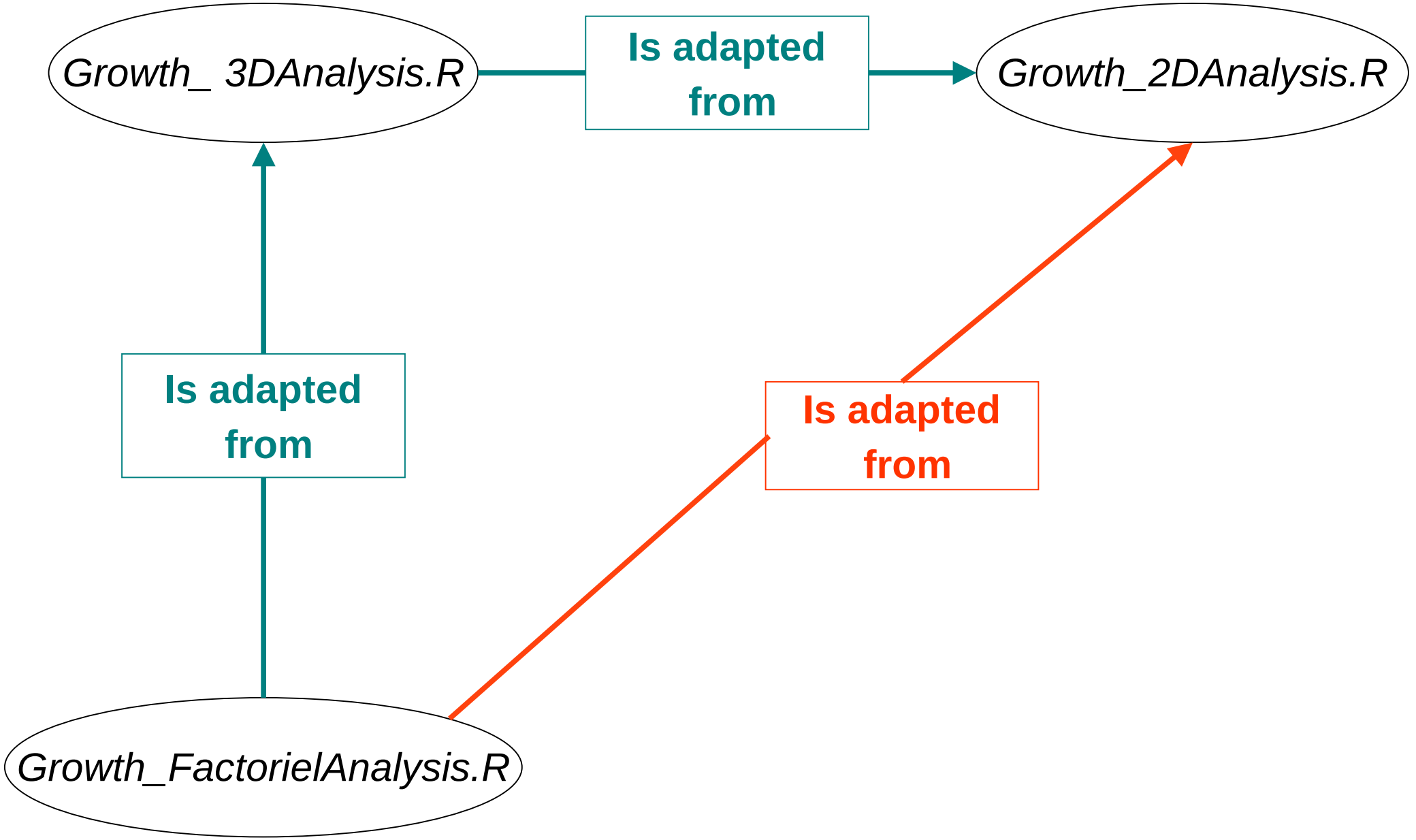
## TRANSITIVITY

```
<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***

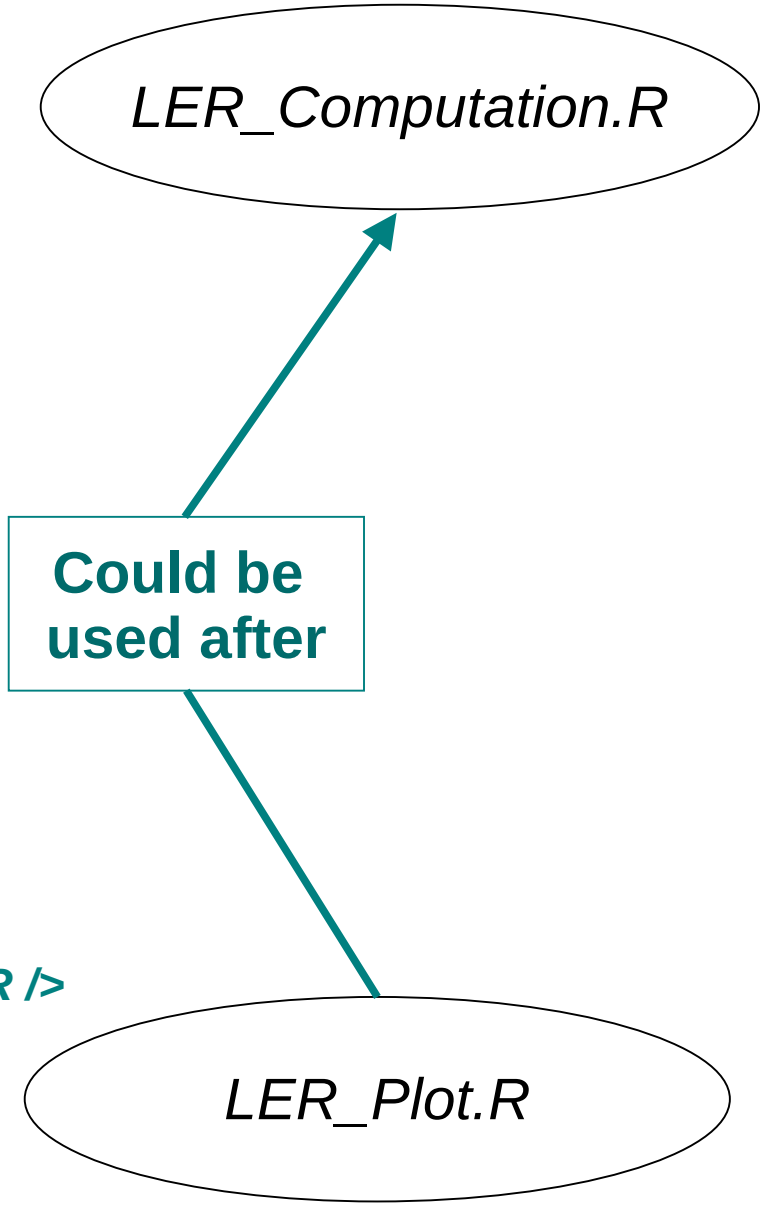
*LER_Computation.R*

# Relations deduced by Inference : *INVERSE OF*

***LER_Computation.R***

***LER_Plot.R***

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

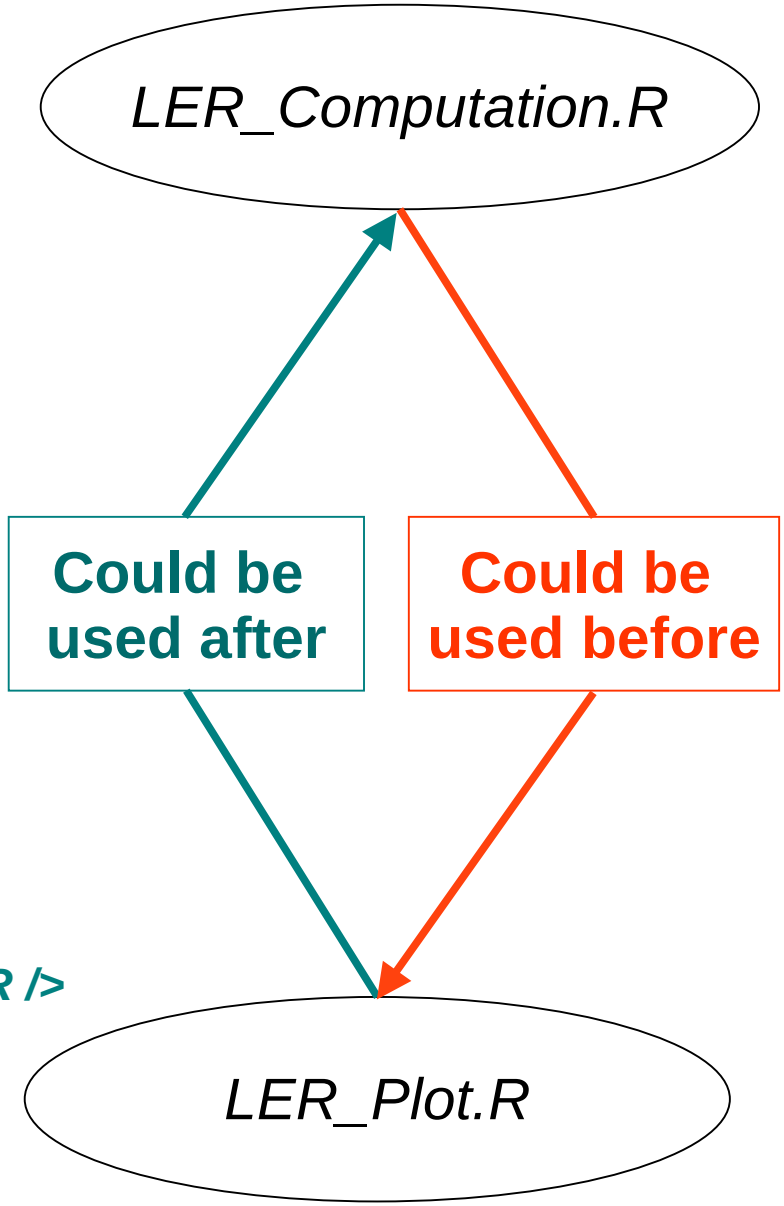


# Relations deduced by Inference : *INVERSE OF*

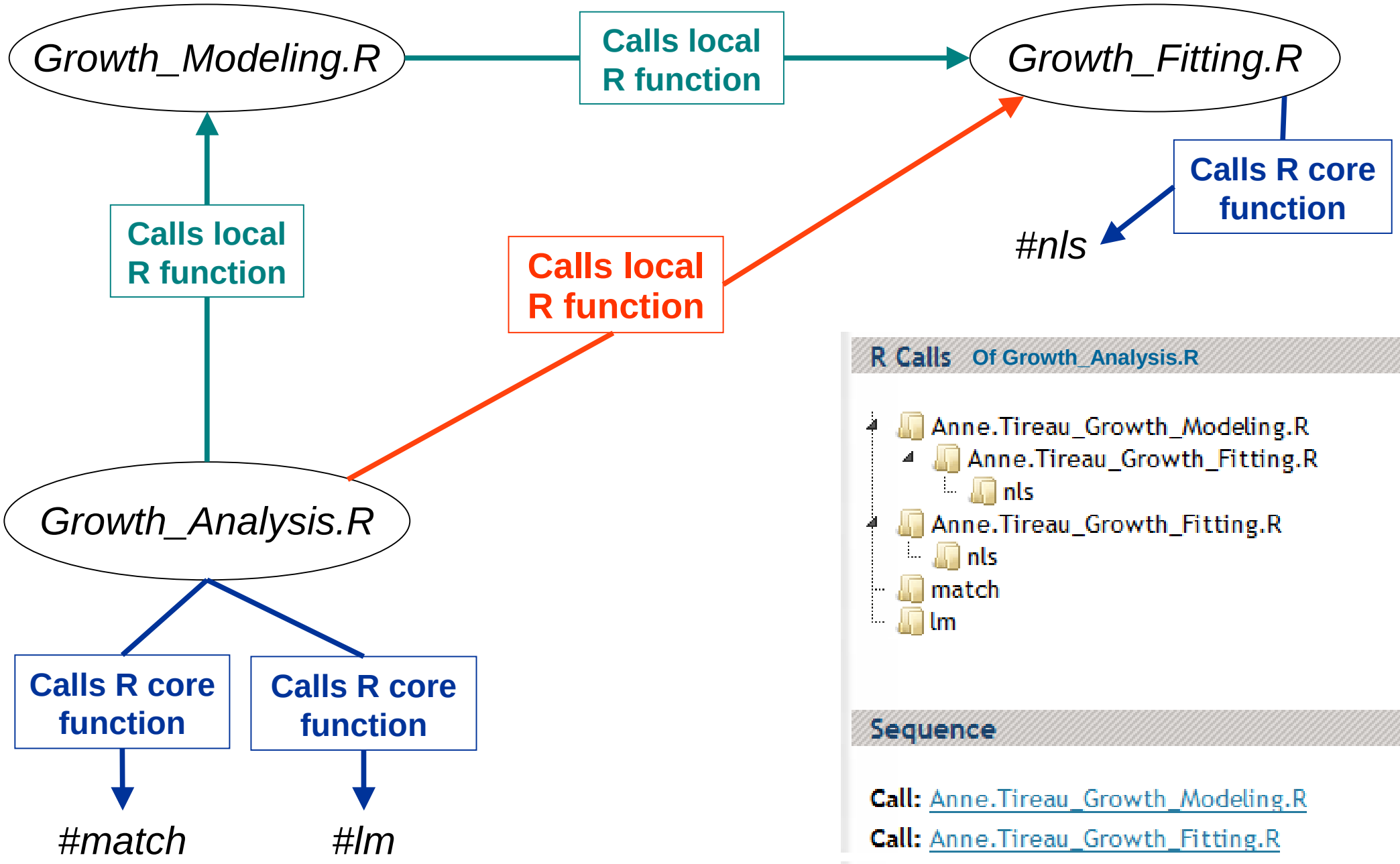
***LER_Computation.R***

***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

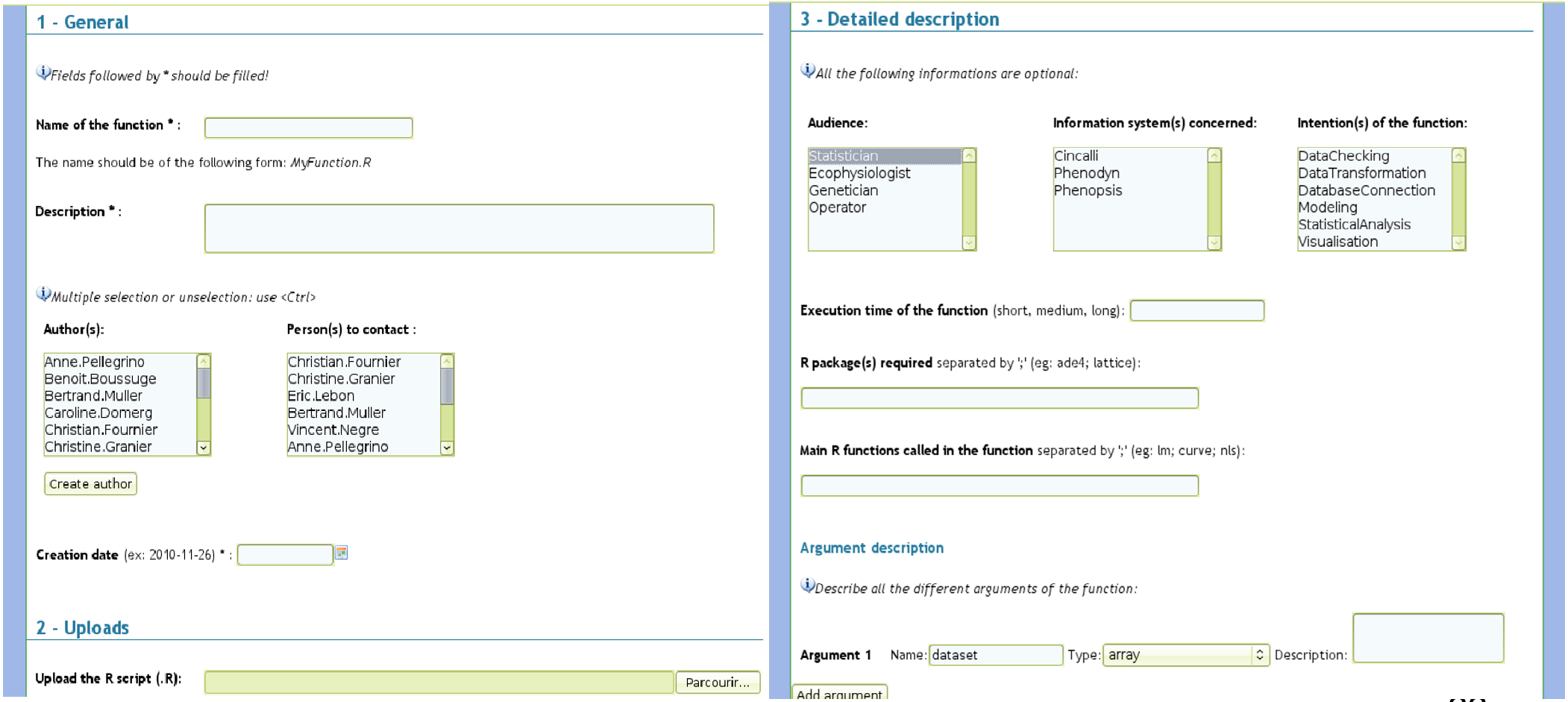
**Thank you**

**Questions ?**

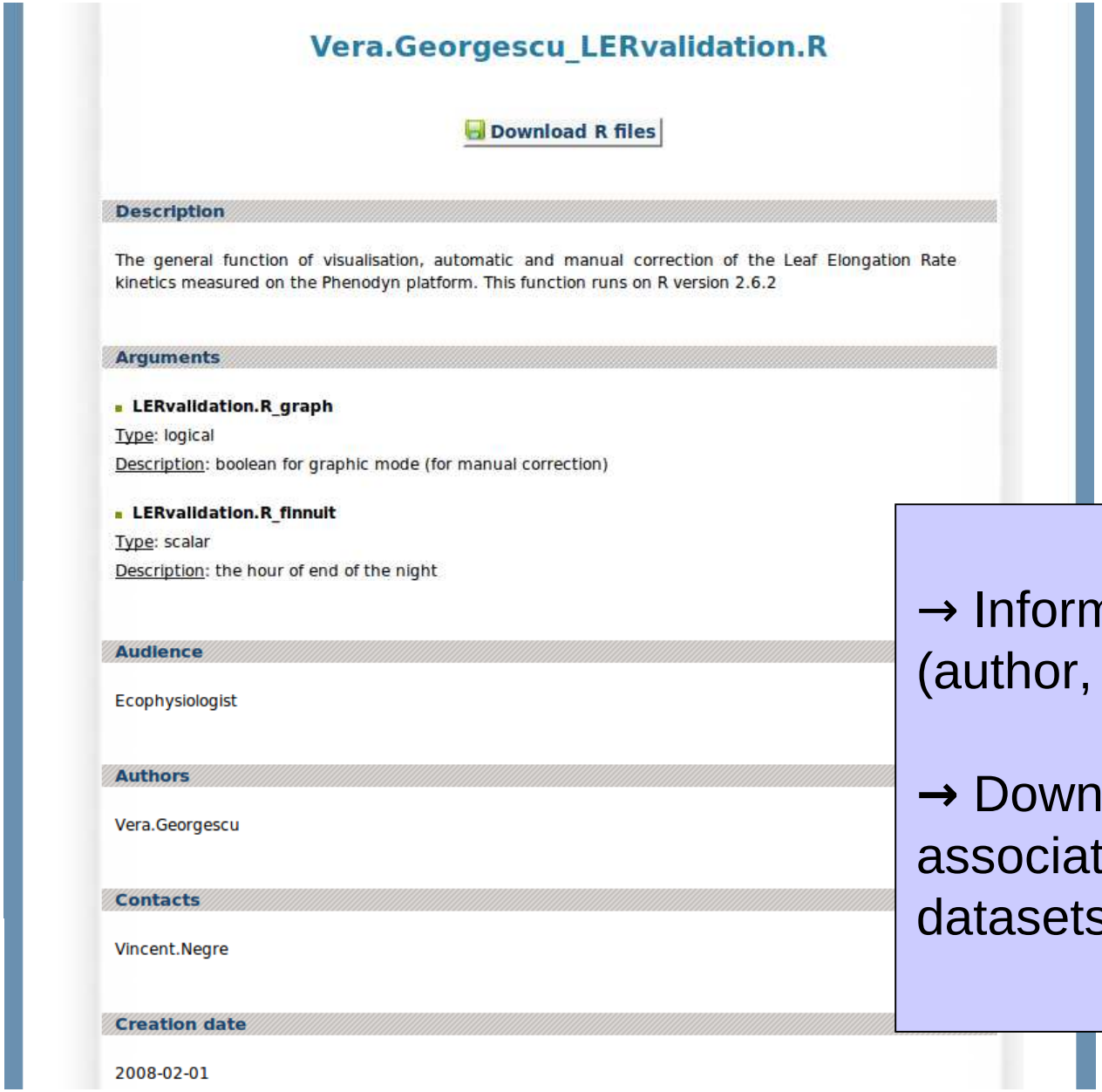
# GUI for edition and creation of annotations

→ A few minutes thanks to pre-filled forms

→ Generation and storage of OWL file



# 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**

Ecophysiologist

**Authors**

Vera.Georgescu

**Contacts**

Vincent.Negre

**Creation date**

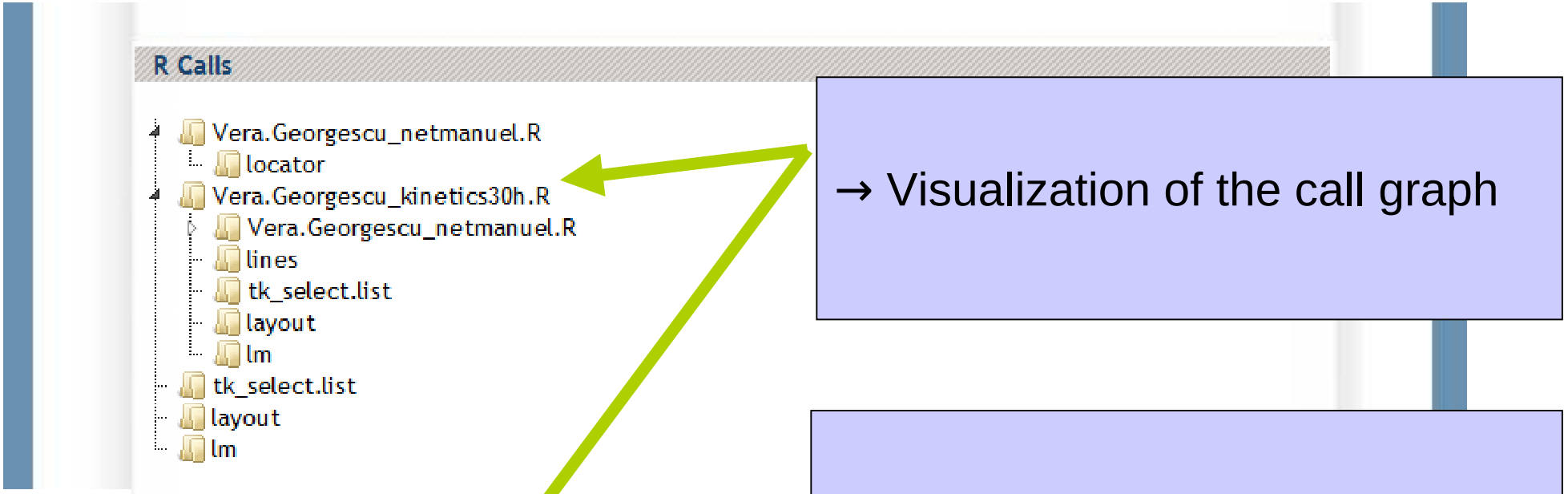
2008-02-01

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

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

# Function consultation card

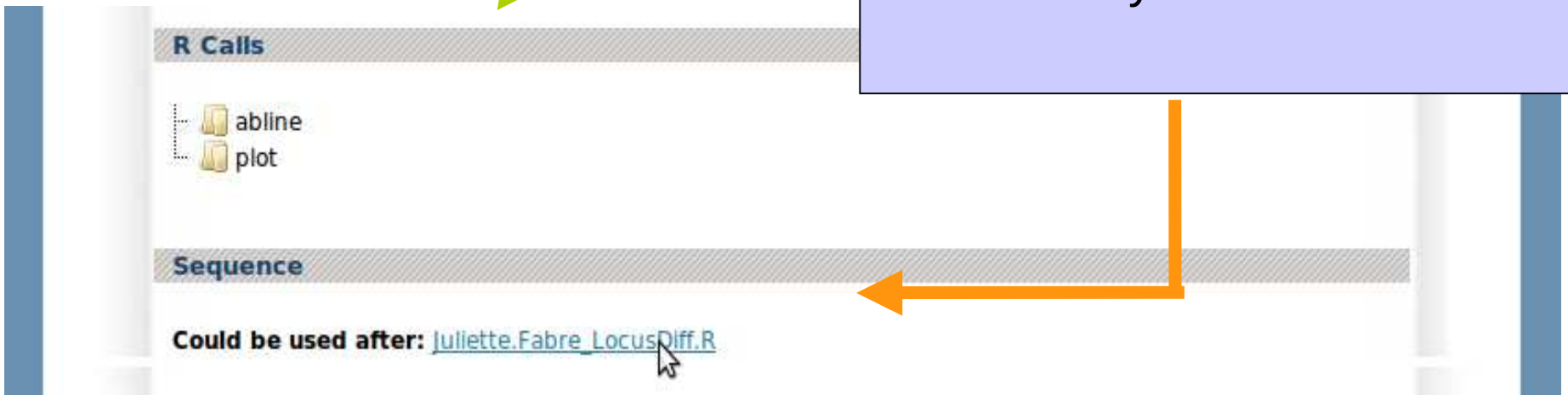
*Example 1*




→ Visualization of the call graph

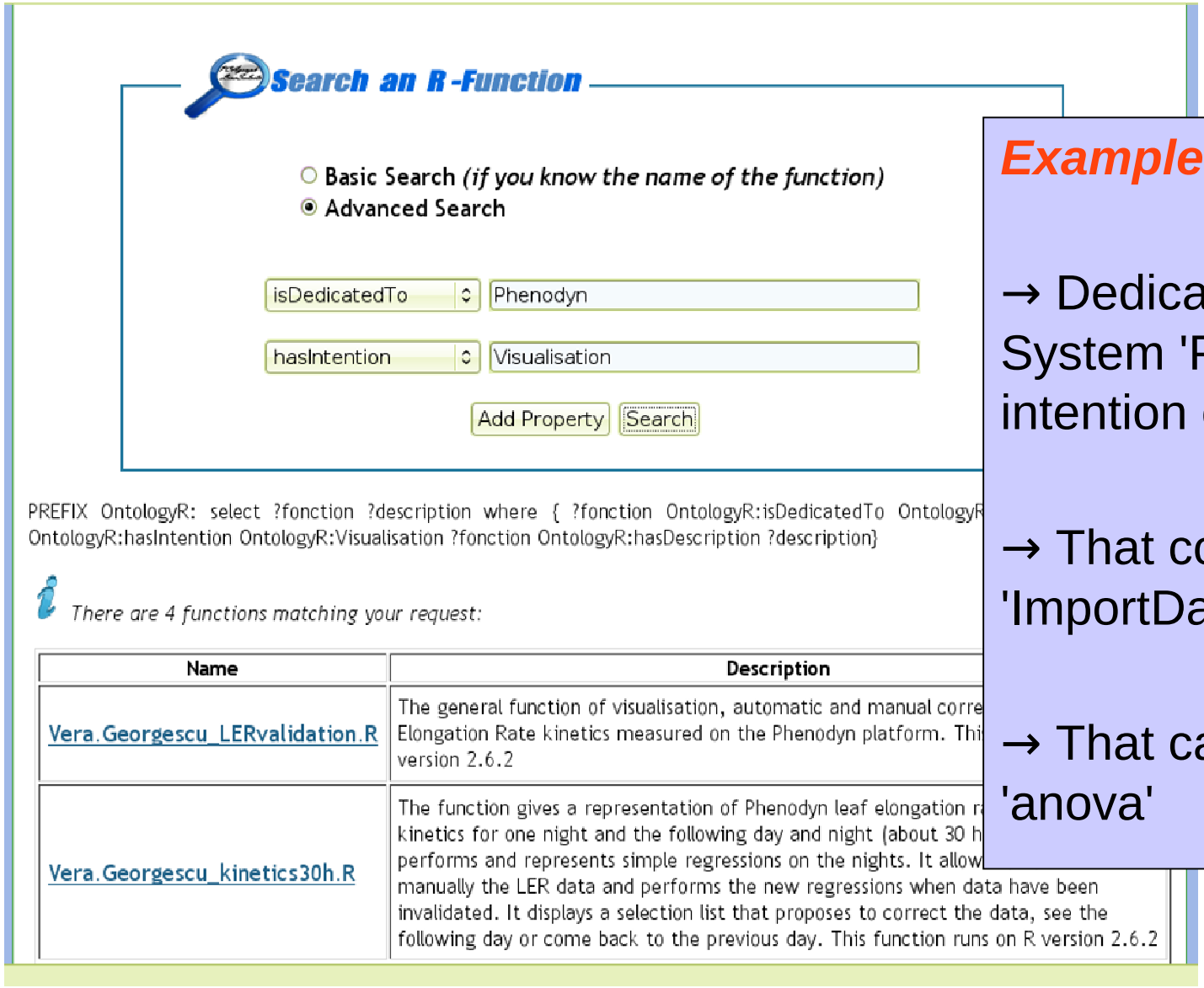
→ Hypertext links towards semantically related functions

*Example 2*



# Powerful search tools

Build a *SPARQL* request adding conditions on the properties




***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'