

SURF

FOUNDATION

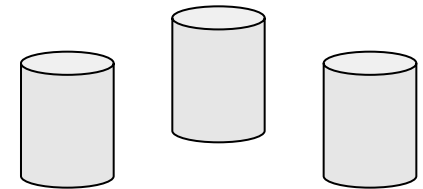
# Statistics on the Usage of REpositories (SURE)

Peter Verhaar, 22 april 2012

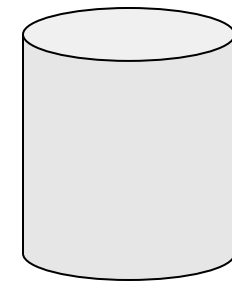
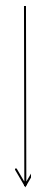


# SURE1

- June 2009 – December 2010
- **Statistics on the Usage of Repositories**
- Infrastructure for the aggregation of usage statistics
- Guidelines

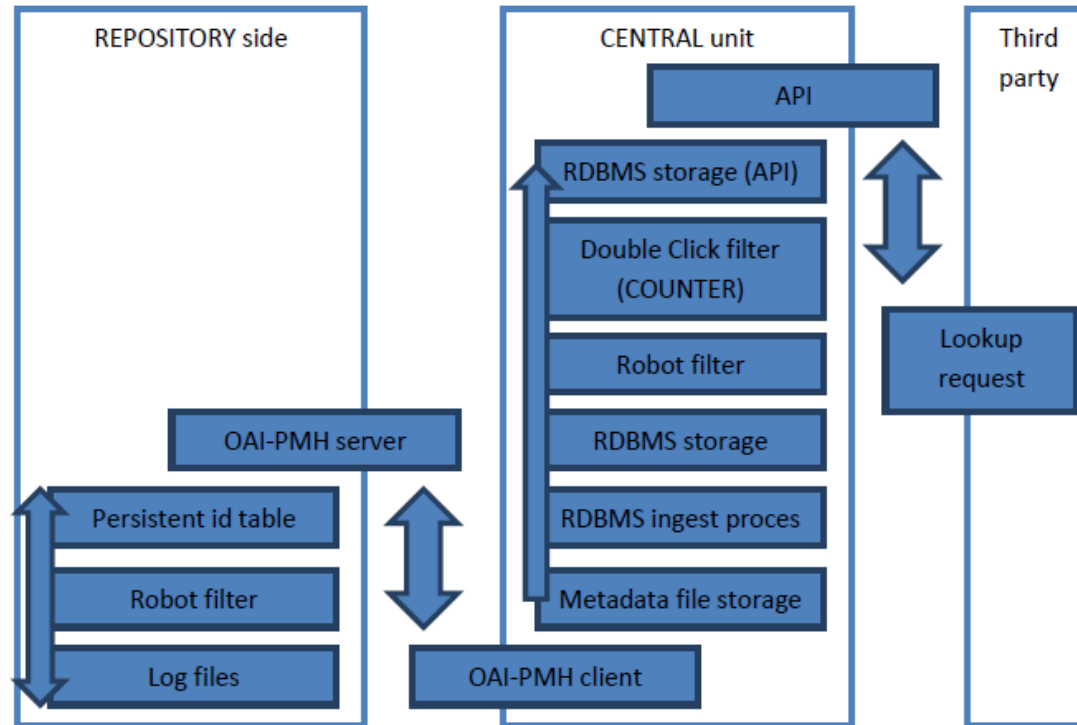


**Repositories**



**Log Aggregator**

# SURE software



- Wiki with project results:  
<http://wiki.surffoundation.nl/display/statistics/Home>



# SURE2

- DANS agreed to manage statistics database
- Objectives of SURE2
  - Stimulate other universities to contribute usage data to the SURE database
  - Build services, mainly for the visualisation of usage metrics

*Data Archiving and Networked Services*

**DANS**



# Dashboards

[BACK TO SEARCH RESULTS](#)

## PUBLICATION

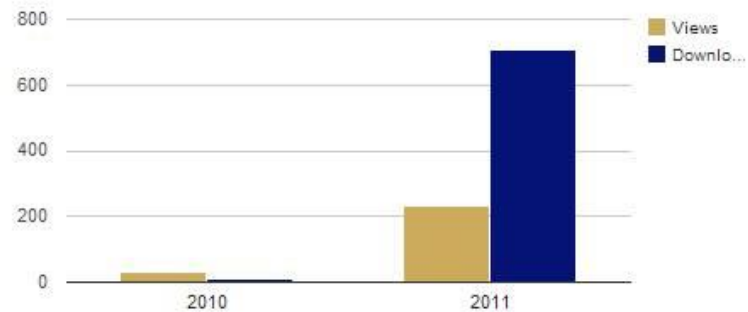
LIZARDS OF BRAZILIAN AMAZONIA (REPTILIA: SQUAMATA) (1995)

[Open access](#)

### Pagina-navigatie:

- [Main](#)
- [Insight Publication](#)

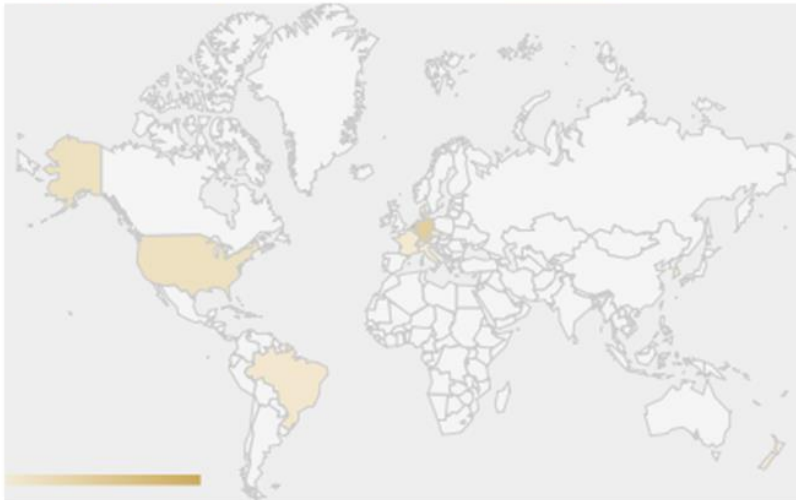
Data available since 2010-10-14





# Countries and referrers

Geographical spread of the visitors [what is this?](#)



Website which refers the visitor to the abstract or full text [what is this?](#)

Referrer	
google scholar	google

# Metrics per author

PERSON  
PROF.DRS. J.G. DE WIT

Main  
Completed research (1)  
Publications (36)  
Metrics

Total number of viewed abstracts and full text documents [what is this?](#)

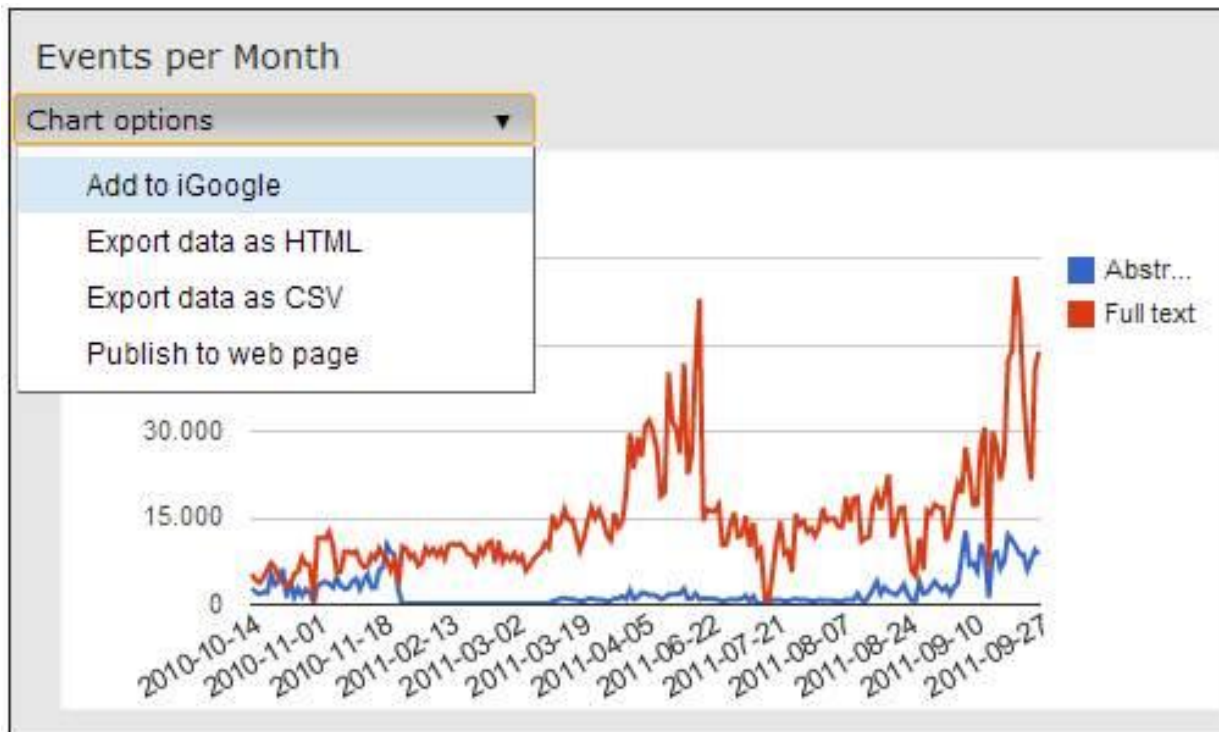
Genre	Total	Abstract	Full Text
article	11	33	289
book	2	6	0
book part	5	8	0
conference paper	1	1	0
report	6	14	0
without metrics	11	0	0

DASHBOARD

from  until    Abstract  Full Text



# Widgets







# A national infrastructure

- Usage data from Erasmus University Rotterdam, NWO/CWI, University of Groningen, TU Delft, University of Maastricht, University of Utrecht, Tilburg University, Wageningen University
- Differences in the configuration of the SURE software undermined integrity and reliability of central data set

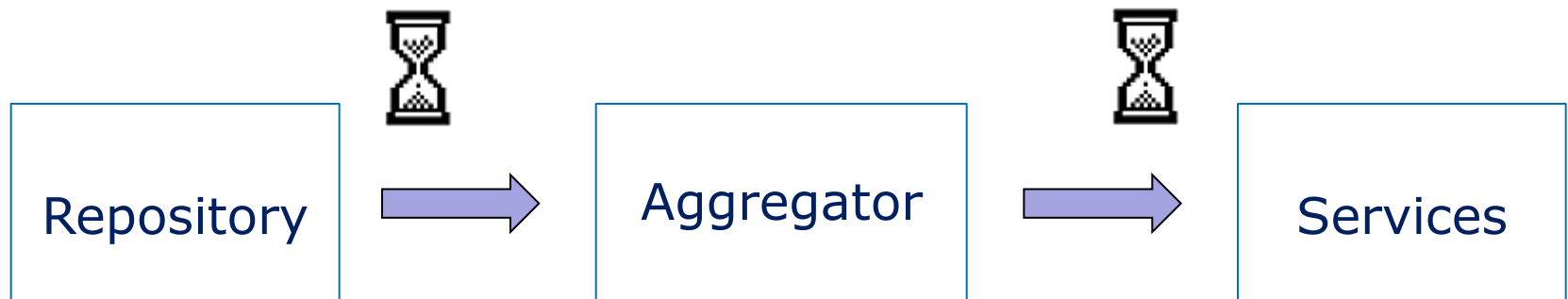


# Technical difficulties

- Errors and inconsistencies
  - Not all repositories could provide PID
  - Metrics on files rather than on publications
  - Metrics on learning objects or student theses
  - Incorrect use of terminology, e.g. 'humanStartPage'

# Technical difficulties

- Performance issues on the central level
  - Harvests take very long
  - Queries on the database take very long



# Alternative scenarios

- Parallel harvesting

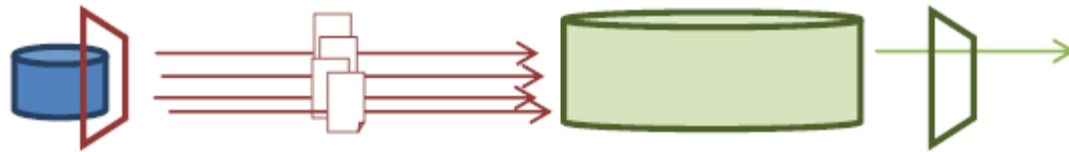


- Distributed databases



# Alternative scenarios

- Tracker code instead of harvests



- No harvesting





# SURE3?

- **Not so SURE ...**