



Citation Statistics in IRStats2

Dr. Martin Brändle

martin.braendle@id.uzh.ch



**Universität
Zürich** UZH

HBZ / Zentrale Informatik



Short Demo



Short Demo

All items > Authors: Hengartner, M O

Available Reports

Summary Citations Deposits Requests

Filter Items

Authors

e.g. Smith, John

Please enter some text to filter by (at least 2 characters)

Dates

Select dates:

Start Date: YYYY/MM/DD



End Date: YYYY/MM/DD



View

-OR- select a range:

Past month



View

Top Cited Items (Scopus)

1.	The biochemistry of apoptosis	Scopus®	4,555
2.	Molecular definitions of cell death subroutines: recommendations of the Nomenclature Committee on Cell Death 2012	Scopus®	643
3.	Programmed cell death: alive and well in the new millennium.	Scopus®	451
	Genetic control of programmed cell		

Top Cited Items (Web of Science)

1.	The biochemistry of apoptosis	Web of Science®	4,212
2.	Molecular definitions of cell death subroutines: recommendations of the Nomenclature Committee on Cell Death 2012	Web of Science®	612
3.	Programmed cell death: alive and well in the new millennium.	Web of Science®	410



Gathering Data (1)

Scopus

Citation count dataset and import plugins v1.1.0, Queensland University of Technology, <http://files.eprints.org/815/> +modifications by UZH

Scopus API allows further use of citation data, given that source is mentioned and a backlink is provided.

Web of Science

WoS Links Article Match Retrieval API, script by Clyde Cerejo (formerly UZH).

Thomson Reuters allowed us further use of citation data. We treat as with Scopus.

Google Scholar

Search configuration Queensland University of Technology, <http://files.eprints.org/815/> +modifications by UZH

Google Scholar does not allow to harvest citation counts.



Gathering Data (2)

Scopus

Daily, in batches of ~2.7K, ~45 min

Updated fields: scopus_impact, scopus_cluster, scopus_datestamp

~41K records with citation data (51%)

Web of Science

Weekly, all live archive of type article, book section, conference item, ~4h

Updated fields: woslamr_times_cited, woslamr_source_url,
woslamr_citing_articles_url, woslamr_related_records_url, doi (if empty)

~32K records with citation data (40%)

IRStats2

Weekly, ~9h



IRStats2 implementation: Processing data

- IRStats2 provides datasets and datatypes
http://wiki.eprints.org/w/IRStats_2_Technical_Documentation
- Predefined datasets: access, eprint, history
- Processor.pm passes on records of the datasets to its associated Stats::Processor modules (e.g. Stats::Processor::Eprint modules)
- Each module provides a specific datum (datatype), and counts/sums associated values
e.g. Stats::Processor::EPrint::DocumentAccess provides
 - datatype doc_status, with evaluated properties
 - fulltext_status = ,full_text' | ,no_full_text'
 - openaccess_status = ,open_access' | ,no_open_access'
- Obtain new datatypes by adapting a Processor module



IRStats2 implementation: Processing data (2)

- Citation statistics: datatype simple: citation count
- From citation database-specific Processor module to a generic value tracker (thanks, Adam!), calling sub-modules for datatype declaration
- Datatypes and associated fields can be specified in the `z_stats.cfg`
- Further abstraction possible?



Display

- Specific Stats::View::ReportHeader and Stats::View::Table plugins
- Replacing standard IRStats2 plugins using plugin mapping