37th IEEE International Conference on Data Engineering
(ICDE 2021), 19-22 April 2021

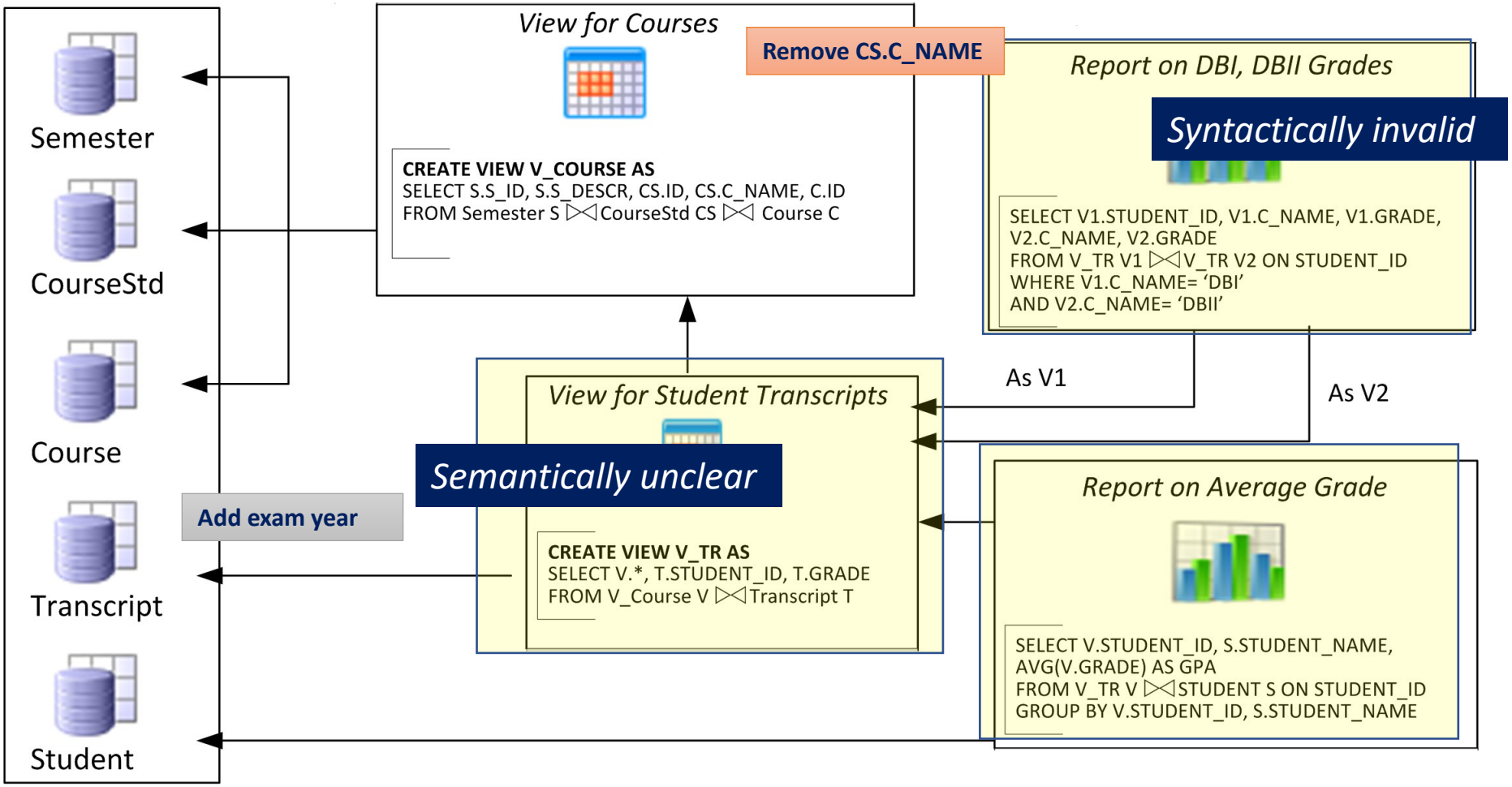
Profiles of Schema Evolution in Free Open Source Software Projects

Panos Vassiliadis

http://www.cs.uoi.gr/~pvassil/projects/schemaBiographies/

Department of Computer Science and Engineering
University of Ioannina, Hellas

Schema Evolution has an impact



The impact can be syntactical (causing crashes), semantic (causing info loss or inconsistencies) and related to the performance

Why is schema evolution so important?

- **Dependency magnets**
  - **Databases are rarely stand-alone**: typically, an entire ecosystem of applications is structured around them
- The impact of change is
  - **Syntactic**: scripts & reports simply crash …
  - **Semantic**: views and applications can become inconsistent or information losing …
  - **… changes in the schema can impact a large (typically, not traced) number of surrounding applications, without explicit identification of the impact & can cause several (parts of) different applications to crash, slow down, or miss data, causing the need for emergency repairing**

Why bother studying schema evolution?

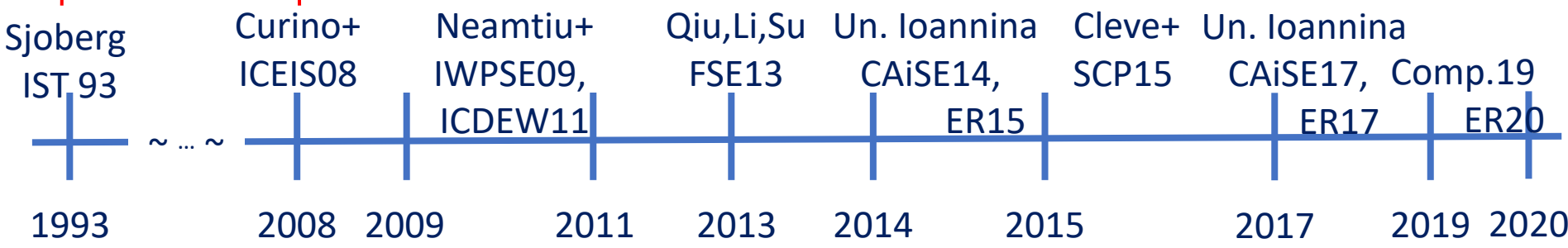
- … to **develop the understanding** of how schemata typically evolve via substantial empirical evidence
  - to identify key **patterns** and features that characterize schema evolution
  - to **quantify** features and characteristics (overall and per pattern)
- … to **contribute to the research community** by clarifying the extent of the presence (absence) of schema evolution in the lives of FOSS projects, and…
  - … to initiate interest on **how to educate** young professionals
  - … to provide evidence for **future research developments** on **how to link schemata to the surrounding code**
- … to allow **managers, curators**, … **assess and predict** the extent of schema evolution in the future of a project, and prepare for it

Timeline of Studies on Relational Schema Evolution

- Historically, nobody from the research community had access + the right to publish to version histories of database schemata
- Open source tools internally hosting databases have changed this landscape, so we are now presented with the opportunity to study the version histories of such "open source databases"

Mind the gap!
(15 years)

Studies on how schemata evolve



NoSQL: Klettke+ BigData17 Scherzinger+ ER20

What are the patterns/laws of schema evolution?

Research Questions

- [RQ1] Is schema evolution extensively present? Is schema evolution a process that frequently encountered, and if yes, to what measurable extent does it occur in terms of frequency and volume?
- [RQ2] Are there consistent patterns in the lives of schemata -- i.e., can we extract families, ("taxa" as in biology) of schemata, with respect to the way they evolve over time?
- [RQ3] What are the quantitative characteristics of schema evolution and how do they perform for different taxa?

Our approach

- Performed the largest study ever: collected 327 schema histories from github
- Isolated 195 out of them that had > 1 commits to the DDL file, cloned the history of project, extracted the history of the DDL file, produced the pairwise differences of subsequent commits of the DDL file, and measured evolution in terms of
  - Heartbeat of changes = time series of change activity
  - Amount of activity, both total and broken down per type
  - Timing characteristics
  - Schema properties (#tables and #attributes)
  - Quantified qualitative characteristics (e.g., spikes of activity)
- For the first time in the literature (to the best of our knowledge), we study this heartbeat of change, and, following an iterative, qualitative process, we grouped projects in taxa of evolution, i.e., families of schemata, which share similar evolution characteristics.

https://github.com/DAINTINESS-Group/

Contributions

- Result generalizability
  - Largest study ever (by one order of magnitude) …
  - Principled collection method …

  => A fairly representative view of Schema Evolution in the FOSS universe (&& not just some hand-picked examples) (with limits)
- Concrete evidence that **although evolution is present, its absence is way more omnipresent**
  - With various implications for us as teachers & researchers
- Identification of **taxa of schema lives** (first time ever)
  - To be used as a forecasting / sw characterization / … tool
- Apart from the **public data, src & results**, a detailed experimental method, nomenclature, visualization and analysis methods to be reused by subsequent studies

Background

Nomenclature, Data Collection, Scope and Threats to Validity

Extraction Process: aimed for massive history collection at very large numbers

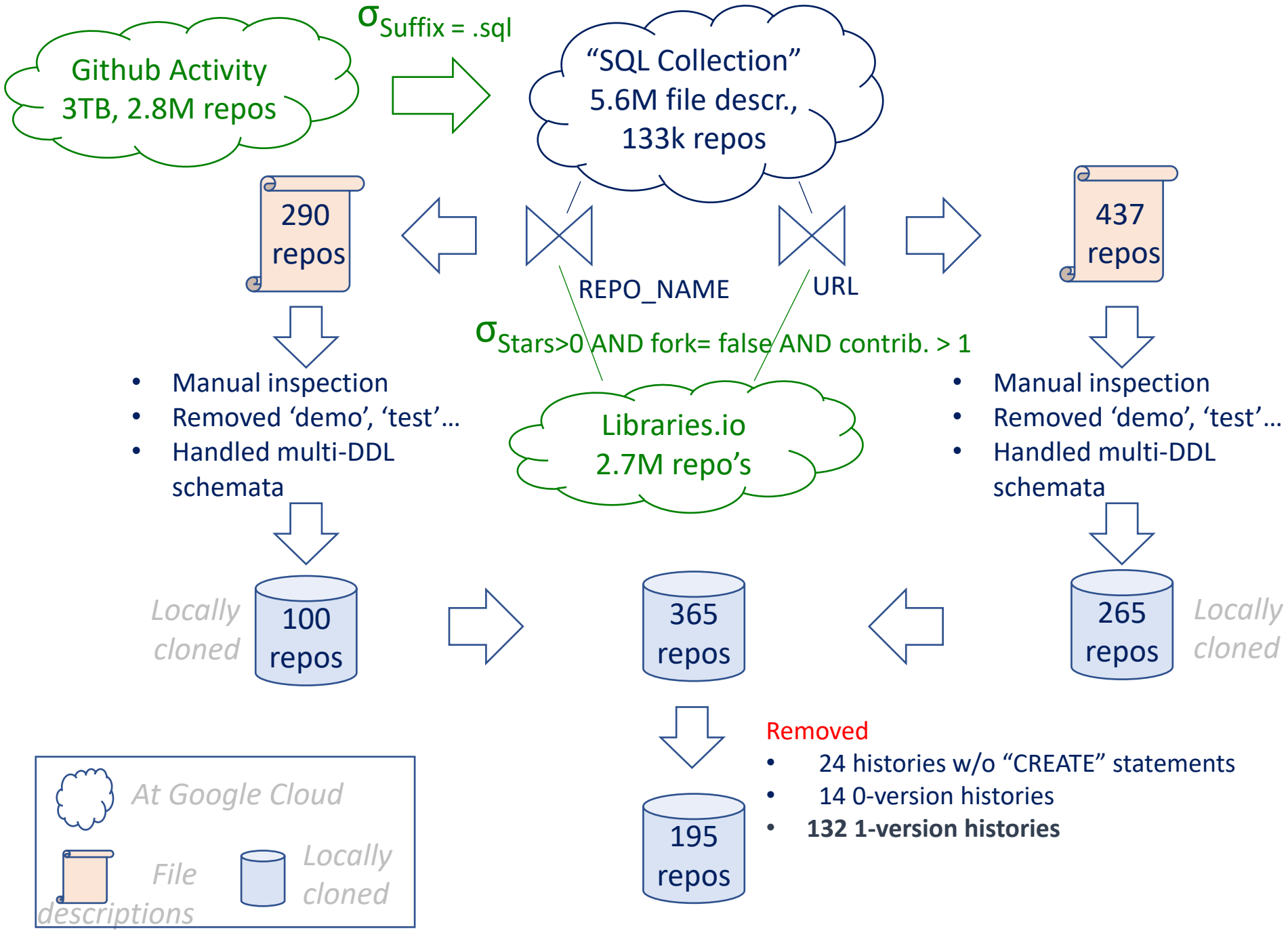
- Queried the GitHub Activity Data dataset from Google Cloud BigQuery (a 3TB+ dataset that contains a full snapshot + the commits of more than 2.8 million open source GitHub repositories) for repos having .sql files → 133K repo's
- Joined this with Libraries.io dataset (metadata for > 2.7M FOSS prj's) and
- filtered for
  - original repositories, with more than 0 stars, more than 1 contributor
- excluding
  - all files with 'test' or 'demo' or 'example' in the path
  - instances of multiple appearances of a DDL file for >1 vendors
  - multiple DDL's (file-per-table mode), incremental maintenance, vendor X language Cartesian Products

=> 365 candidates, locally cloned, cleaned from empty .git, .sql files with no CREATE TABLE statements, …, which eventually led to the final data collection.

=> RESULT: 327 histories out of which

… 132 (40%) with just a single commit ⇔never changed ANYTHING)!!!

… 195 histories with at least an extra commit, which we subsequently used


Github Activity
3TB, 2.8M repos
σSuffix = .sql
"SQL Collection"
5.6M file descr.,
133k repos
290
repos
437
repos
REPO_NAME
URL
σStars>0 AND fork= false AND contrib. > 1
Libraries.io
2.7M repo's
• Manual inspection
• Removed 'demo', 'test'…
• Handled multi-DDL schemata
• Manual inspection
• Removed 'demo', 'test'…
• Handled multi-DDL schemata
Locally cloned
100
repos
365
repos
265
repos
Locally cloned
195
repos
Removed
• 24 histories w/o "CREATE" statements
• 14 0-version histories
• 132 1-version histories
At Google Cloud
File descriptions
Locally cloned

We work with significant projects

- In whatever follows, remember that we have not selected just any random project, but rather,…
- we intentionally restricted our scope to original, stared projects, where people were actually contributing effort to develop and maintain.
- Overall, 65% of projects spanned more than 24 months and 77% more than a year.

Post-identification workflow for each of the 195 projects

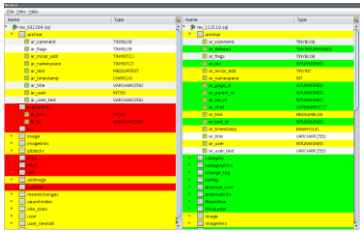
Locally cloned repo

Project's cloned repo from GitHub

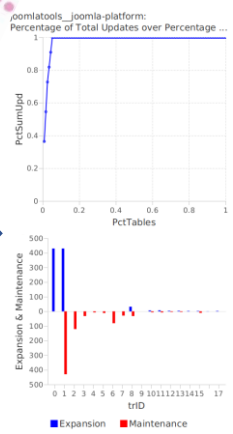
Hecate: SQL schema diff extractor

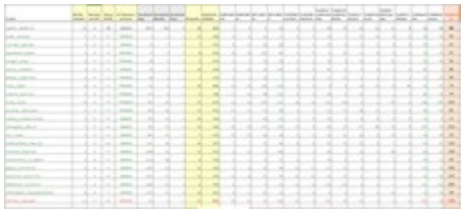
History

Schema History = sequence of DDL committed snapshots



Heraclitus Fire: a chart/stats extractor



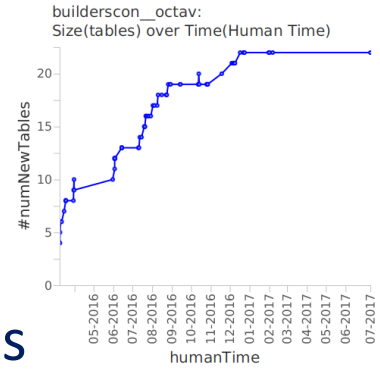


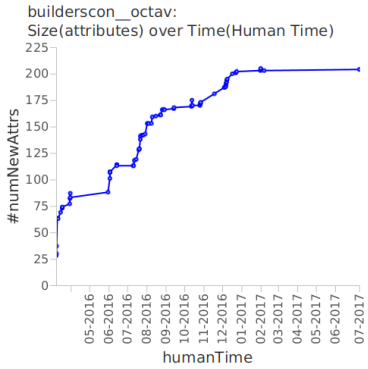
https://github.com/DAINTINESS-Group/Hecate

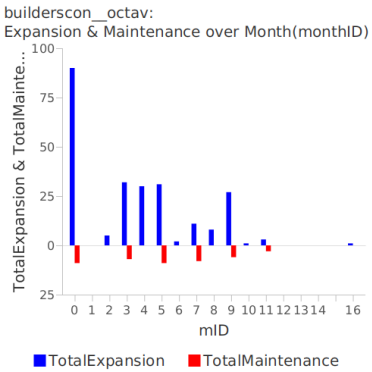
https://github.com/pvassil/HeraclitusFire

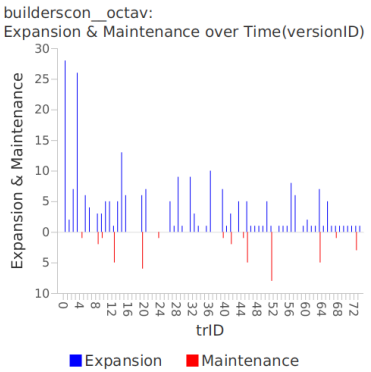
https://github.com/DAINTINESS-Group/Schema_Evolution_Datasets/

Eventually, for each project, we ended up with the automatically extracted
+ time series of changes
+ collected stats on timing, schema size & activity
+ extra statistics manually extracted







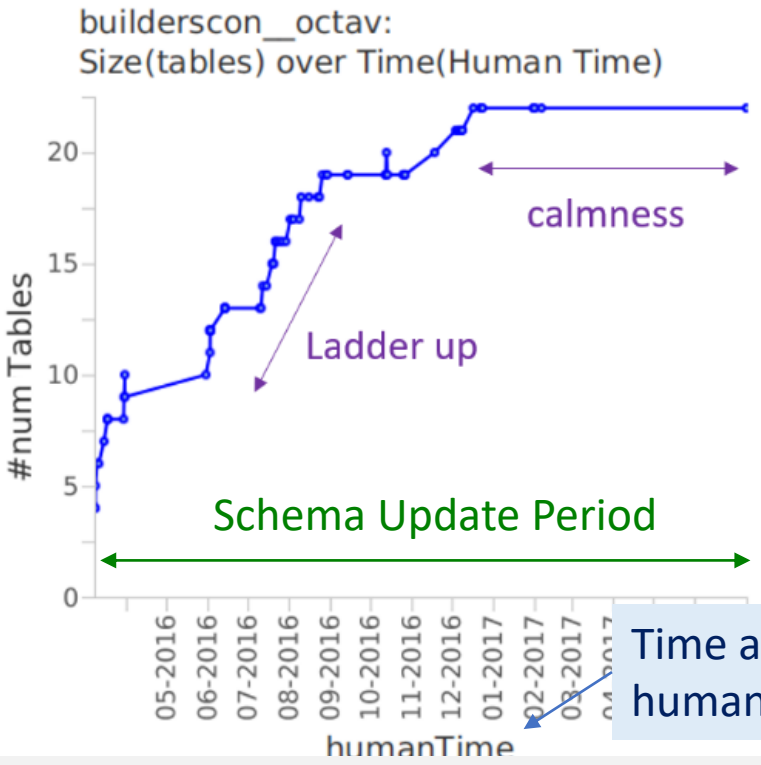


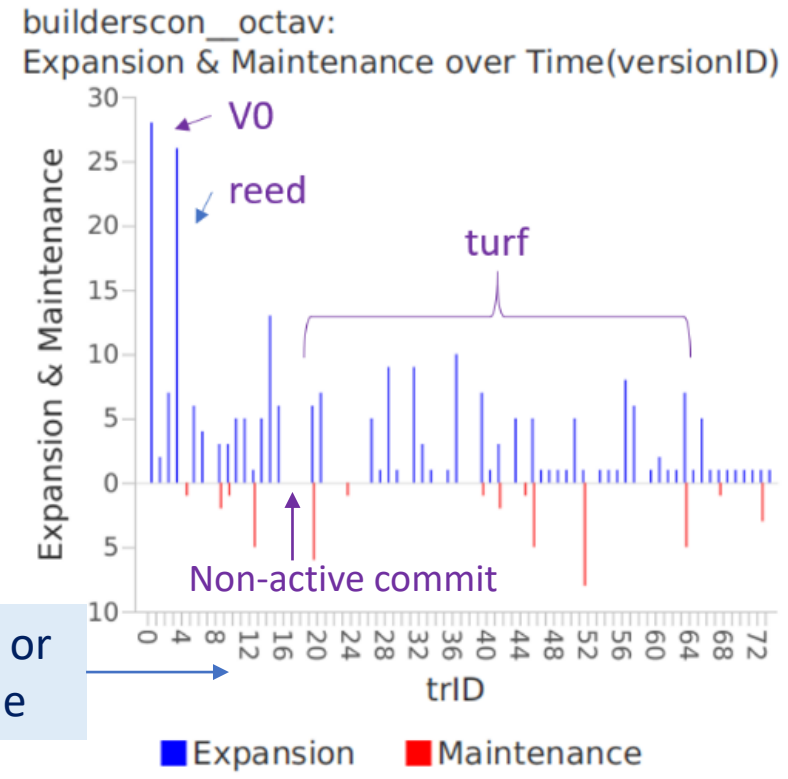
Project	#Active commits	#Areeds postV0	#ATurf postV0	Turf absence / presence	DurationIn Days	DurationIn Months	DurationIn Years	#Commits	ActiveCommitRatio	#Tables@Start	#Tables@End	#Attrs@Start	#Attrs@End	TotalTableInsertions	TotalTableDeletions	TotalAttrInsWithTableIns	TotalAttrbDelWithTableDel	TotalAttrInjected	TotalAttrEjected	TatalAttrWithTypeUpd	TotalAttrInPKUpd	TotalExpansion	TotalMaintenance	TotalActivity
nawork__nawork-uri	22	2	20	ABSENCE	3023	100	9	26	85%	1	4	2	23	4	1	30	13	19	15	11	0	49	39	88
jasdel__harvester	7	2	5	PRESENCE	4	1	1	12	58%	4	6	14	17	4	2	14	7	15	19	0	0	29	26	55
curt-labs__GoSurvey	4	1	3	ABSENCE	15	1	1	7	57%	9	11	63	86	2	0	22	0	4	3	8	0	26	11	37
spaceboats__busbus	5	2	3	PRESENCE	17	1	1	8	63%	14	10	107	82	1	5	3	31	11	8	0	4	14	43	57
milogert__ocdns	6	1	5	PRESENCE	38	2	1	9	67%	3	5	10	28	4	2	19	9	10	2	15	0	29	26	55
joyplus__p2oadmin	4	2	2	ABSENCE	43	2	1	10	40%	2	13	18	103	11	0	85	0	0	0	1	0	85	1	86
keybase__node-client	4	1	3	PRESENCE	60	2	1	7	57%	2	2	16	7	3	3	11	21	2	1	1	0	13	23	36
h2oai__steam	8	1	7	PRESENCE	62	3	1	10	80%	24	24	128	128	0	0	0	0	10	10	14	38	10	62	72
CityGrid__twonicorn	7	2	5	PRESENCE	174	6	1	9	78%	17	23	80	131	6	0	46	0	14	9	1	0	60	10	70
hurad__hurad	10	1	9	PRESENCE	287	10	1	15	67%	12	13	105	121	16	15	142	129	4	1	39	0	146	169	315
pw-press__web-project	5	2	3	PRESENCE	293	10	1	11	45%	11	11	72	69	4	4	12	21	18	12	2	0	30	35	65
lamassu__lamassu-scripts	5	1	4	ABSENCE	754	25	3	16	31%	5	7	26	41	2	0	15	0	6	6	0	0	21	6	27
TalkingData__OWL-v3	7	1	6	ABSENCE	792	27	3	14	50%	26	33	171	245	15	8	84	22	51	39	18	0	135	79	214
n2n__rocket	4	1	3	ABSENCE	801	27	3	7	57%	10	10	46	49	5	5	23	21	10	9	10	0	33	40	73
brettkromkamp__topic_db	9	1	8	ABSENCE	821	28	3	16	56%	7	7	37	40	8	8	42	42	12	9	7	4	54	62	116
kronusme__dota2-api	7	2	5	ABSENCE	1008	34	3	15	47%	6	13	74	169	7	0	85	0	13	3	48	0	98	51	149
TwitchScience__rs_ingester	6	1	5	ABSENCE	1126	38	4	8	75%	2	4	9	17	4	2	16	9	4	3	1	0	20	13	33
gugoan__economizzer	6	2	4	ABSENCE	1285	43	4	12	50%	3	8	18	66	16	11	117	74	6	1	1	0	123	76	199
anchorcms__anchor-cms	7	1	6	ABSENCE	1524	51	5	17	41%	10	10	57	59	10	10	57	57	5	3	3	0	62	63	125
GoBelieveIO__im_service	7	2	5	ABSENCE	1586	53	5	9	78%	2	2	5	13	14	14	134	134	8	0	7	0	142	141	283
alextselegidis__easyappointments	8	1	7	ABSENCE	1716	57	5	19	42%	9	11	61	73	2	0	8	0	5	1	21	1	13	23	36
EPICPaaS__appmsgsrv	21	3	18	PRESENCE	63	3	1	24	88%	10	23	81	184	17	4	125	32	14	4	3	0	139	39	178

- Project
- #Active commits
- #Areeds postV0
- #ATurf postV0
- Turf Ratio
- Turf absence / presence
- DurationInDays
- DurationInMonths
- DurationInYears
- #Commits
- #Tables@Start
- #Tables@End
- #Attrs@Start
- #Attrs@End
- TotalTableInsertions
- TotalTableDeletions
- TotalAttrInsWithTableIns
- TotalAttrbDelWithTableDel
- TotalAttrInjected
- TotalAttrEjected
- TatalAttrWithTypeUpd
- TotalAttrInPKUpd
- TotalExpansion
- TotalMaintenance
- TotalActivity

Nomenclature

Heartbeat: changes @each commit
H = {commit, {change}*}*





Time as: id or human time

Commits: non-active vs active
Active: reed(TotalActivity > 14 attr's)
turf (otherwise)

Schema Update Period: time span between 0^{th} (originating v.) and last commit for schema updates

Project Update Period: resp., for all project updates

Expansion: attr's born with new table, injected to existing tables

Maintenance: att's deleted with deleted table, ejected from surviving table, data type change, PK change

Total activity = Expansion + Maintenance

The unit of measurement is **#affected attributes**

Scope of the study

- We are interested in the monitoring of the **evolution** of the logical-level relational **schema** for significant Free Open Source Software projects, hosted in GitHub.
- We are not covering or generalizing to
  - … proprietary schemata outside the FoSS domain,
  - … conceptual or physical schemata,
  - … non-relational schemata, e.g., XML, JSON, …

Threats to validity: External Validity

- External validity: we argue that our collection is a very good representative of significant FOSS projects
  - GitHub is the main public repository for FoSS prj's.
  - We applied the filter of more than one contributor, more than 0 stars and non-forking
  - Subsequently, filtered out tests, examples & demos
  - Project domains include Content Management Systems, IoT Management on the cloud, Task Management Systems for O/S's, Messaging Platforms, Systems for the management of Scientific Data, Web on-line stores, On-line Charging Systems (OCS)…
- Limitations:
  - Multi-vendor DDL: covered only one vendor
  - Non-sql schemata, non .sql suffixes, multi-file DDL, incremental definitions of DDL

Threats to validity: Experimental Reliability

- We tested our extraction scripts with OpenCart, the largest of our studied projects for which we had a previous past extraction of its history, in 2016.
  - almost identical result, as only one commit out of 412 was missing from the GitHub history we extracted.
- 100% match for manual test of the histories of the retrieved files for a random sample of 50 cases.
- 100% match for removed projects from GitHub at the time of the cloning via a sample of 7 of them.
- Concerning our own software, we did extensive checks to our metrics computation tools.

The Taxa of Schema Evolution

Can we extract profiles of schema evolution?

Taxa extracted via an iterative, manual, qualitative process

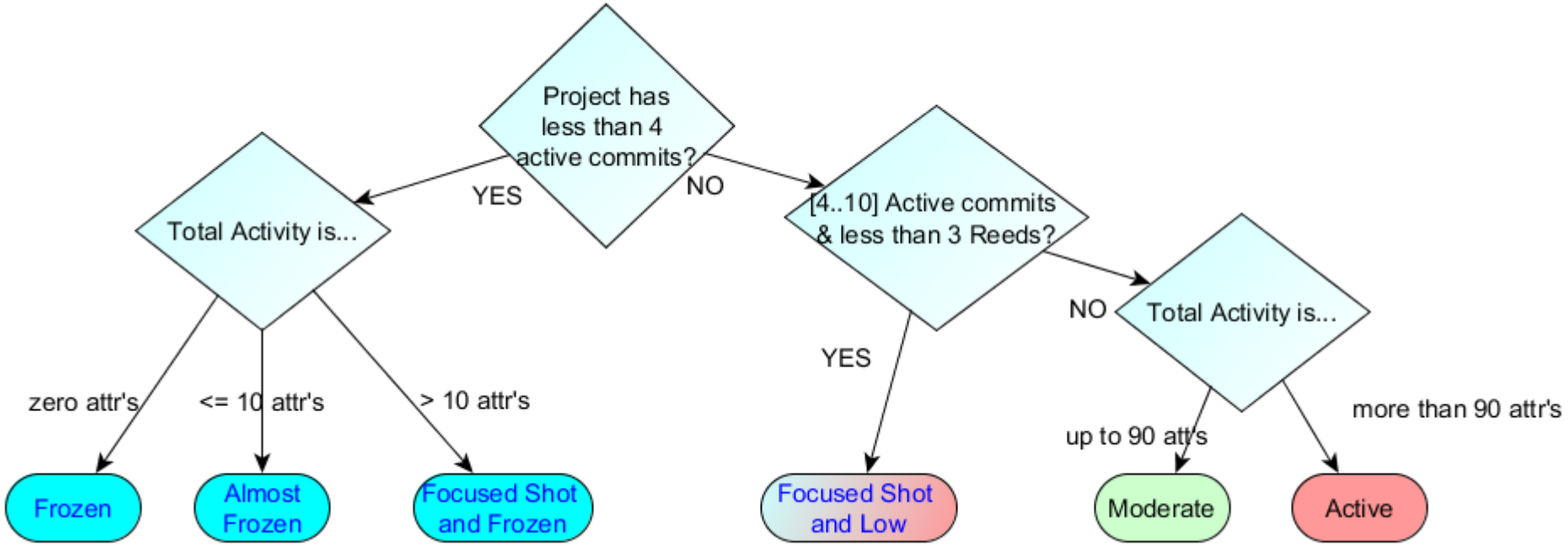
- Given the histories and the charts and statistics extracted by our software, iteratively, manually, and qualitatively grouped the projects …
- … into different groups with similar profile of evolutionary behavior
- … to which we refer to as *taxa*

- Later, a classification tree of taxa was produced to summarize our findings
- A Kruskal-Wallis analysis verifies the difference of the taxa in terms of active commits and total activity

The taxa of schema evolution are:

1. completely **frozen** schema histories with zero change at the logical level;
2. **almost frozen** histories of very small change, typically with few intra-table attribute modifications;
3. almost frozen histories but with a single spike of change and almost no other change (**Focused Shot and Frozen**);
4. histories of **moderate evolution**, without spectacular changes, but rather small deltas spread throughout the life of a project;
5. projects with evolution similar to the moderate one but also with a pair of spikes on their activity (**Focused Shot and Low**);
6. histories of **active** projects, typically with significant amount of change both as intra-table change and in terms of table generation and eviction.

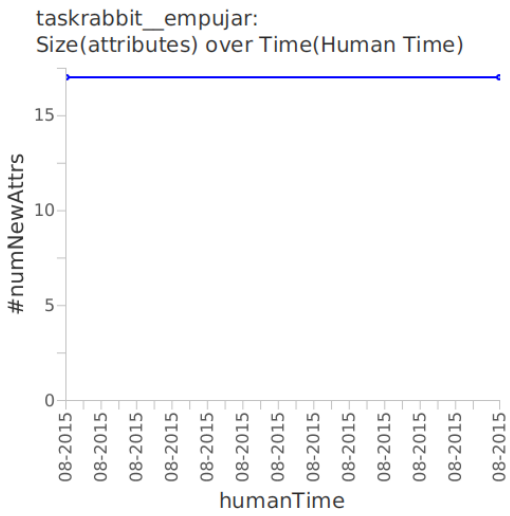
Taxa of Schema Evolution for FOSS Projects

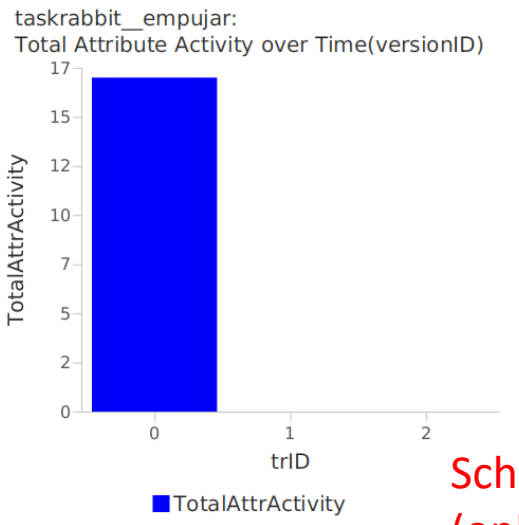


Once the manual, iterative process of taxa was completed, it was also possible to provide a classification scheme…

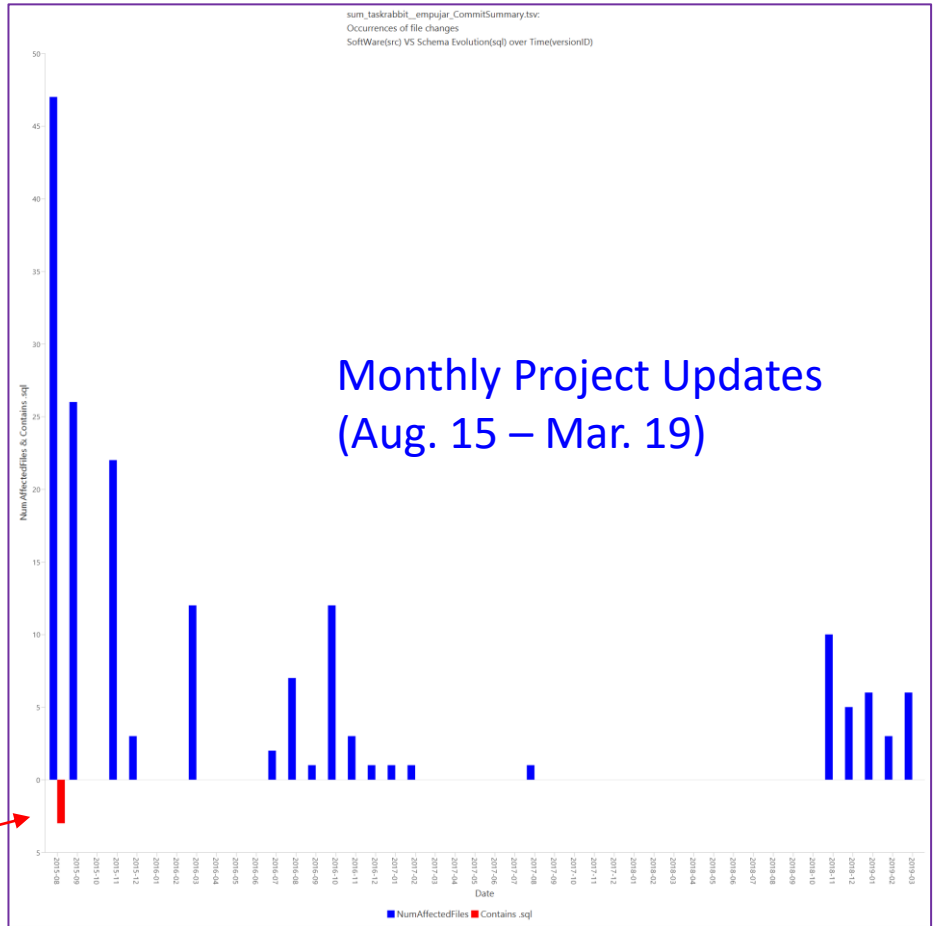
Frozen

34 projects out of 195 came with more than one commits, but with zero changes in their logical schema





Attn: these are NOT dead or toy projects!

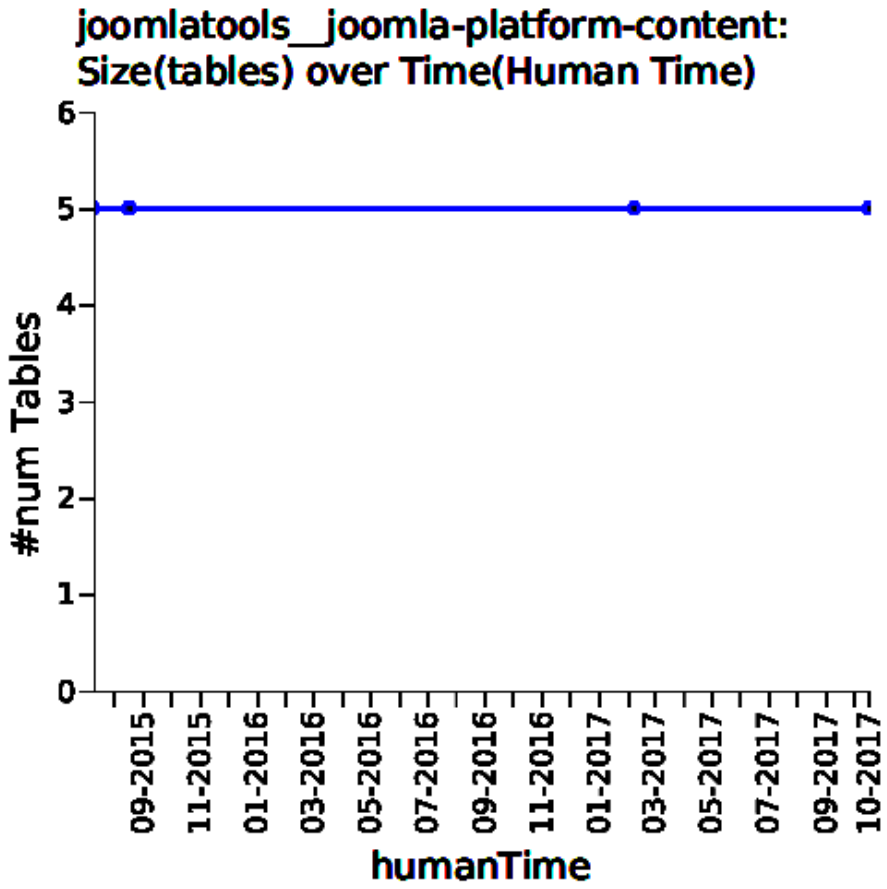


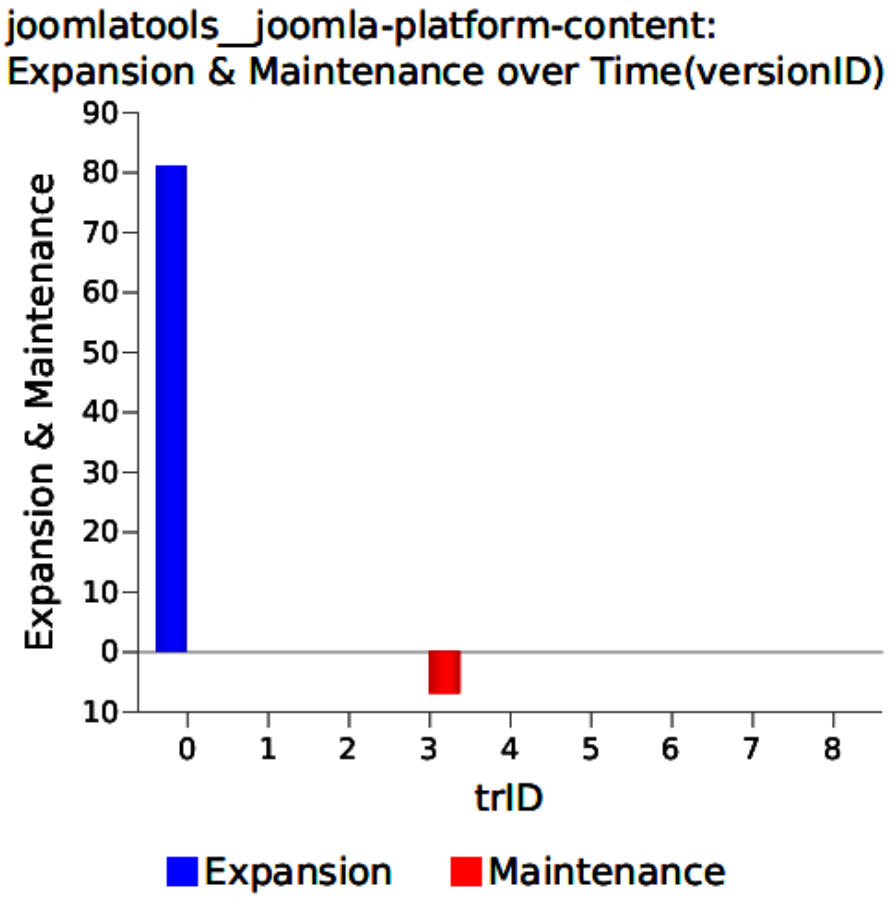
Monthly Project Updates
(Aug. 15 – Mar. 19)

Schema Updates
(only in 1st month)

Almost Frozen

65 of the 195 projects, came with at most 3 active commits and change less or equal to 10 updated attributes
(75% of them with a flat sch. Line)

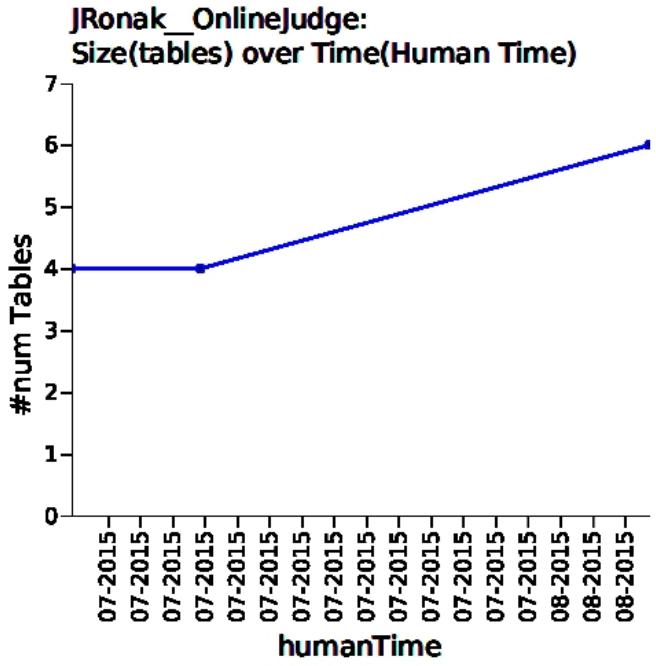


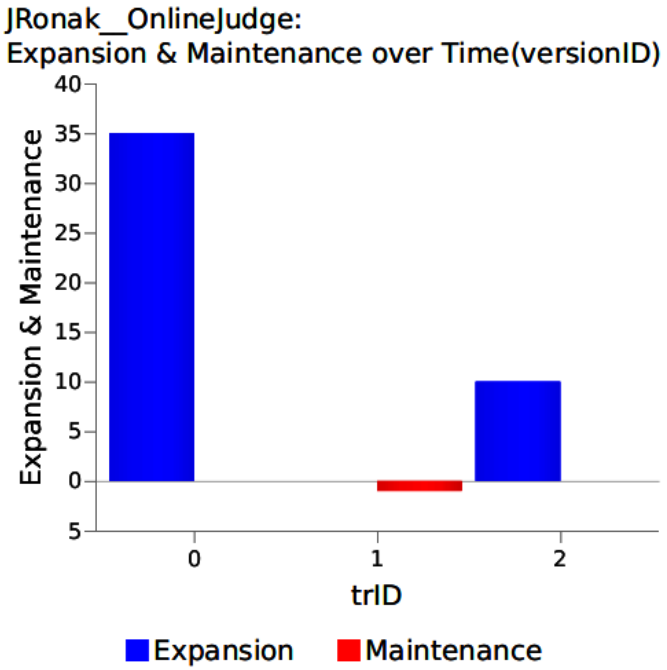


Focused & Frozen

25 projects based around one or two schema modifying transitions, (often a single reed)

- In 36% of them, just attr. injections
- In 52% of them, just a single step-up the schema line





Frozen & Almost Frozen: the absence of evolution is much more evident than its presence (and, yes, this is bad news)

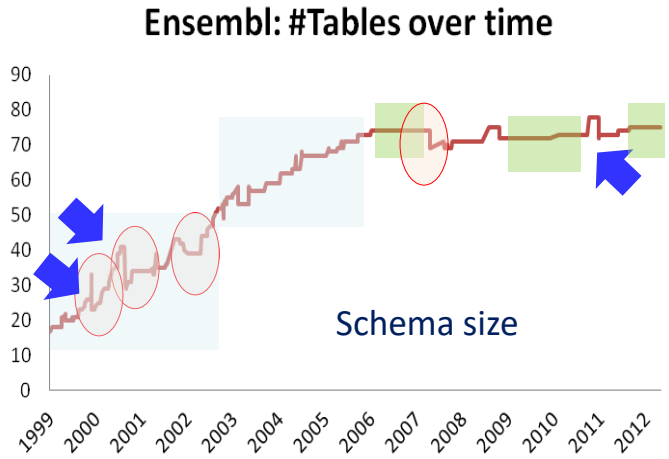
- 70% of the projects, demonstrated total absence or very small presence of change.
- Out of the 327 repositories that we cloned,
  - 132 (40%) had a single commit for their schema (i.e., no change) whatsoever,
  - 34 (10%) had more than 1 commits, but zero changes at the logical-level schema, and,
  - 65 (20%) were almost frozen (with less than 4 active commits and 10 modified attributes).

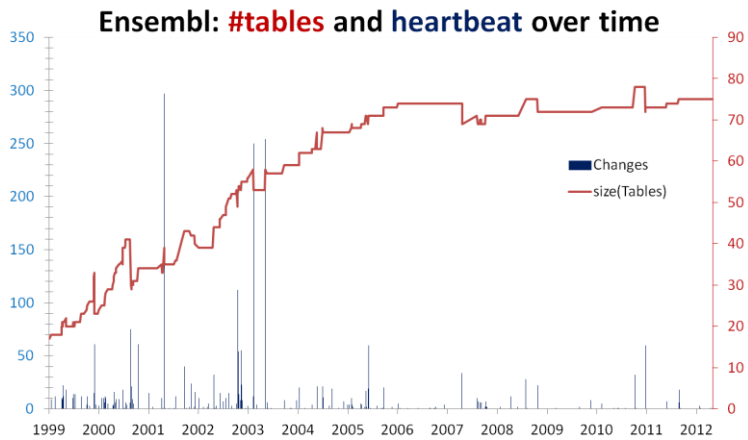
Within the 195 that we cloned

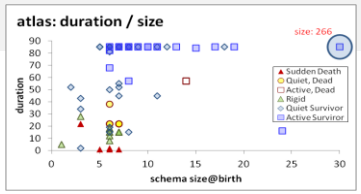
- We have called this phenomenon gravitation to rigidity in our past research [IS15, IS17, JoDS17]

http://www.cs.uoi.gr/~pvassil/projects/schemaBiographies/

Gravitation to Rigidity: the reluctance to evolve the schema is omnipresent, stronger than the tendency to evolve, and grows stronger over time!

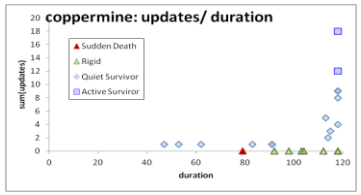




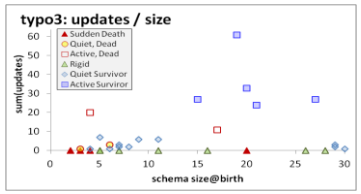


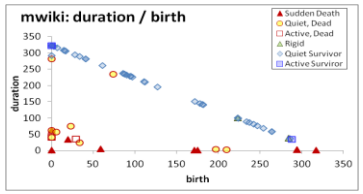
Γ

Individual Tables

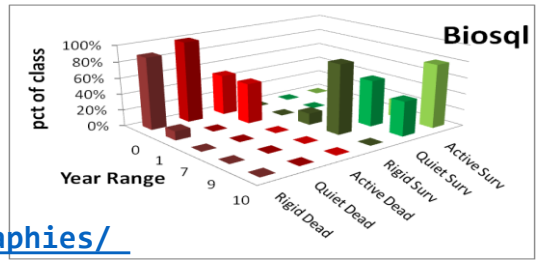


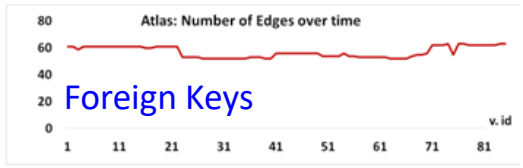
J

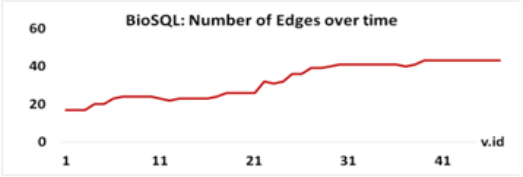


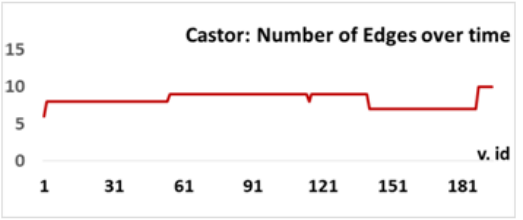


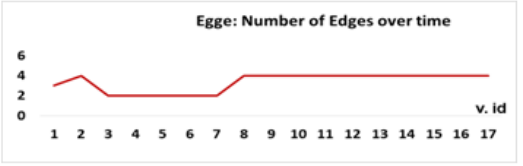
void

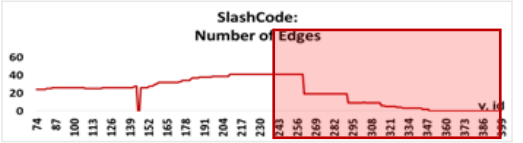


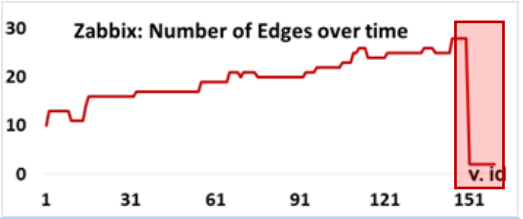












To probe further (code, data, details, presentations, …)
www.cs.uoi.gr/~pvassil/projects/schemaBiographies/

People try very hard NOT to change the schema…

"In a survey of 20 database administrators (DBAs) at three large companies in the Boston area, we found that … DBAs try very hard not to change the schema when business conditions change, preferring to "make things work" without schema changes."

M. Stonebraker, R. C. Fernandez, D. Deng, and M. L. Brodie, "Database decay and what to do about it," Commun. ACM, vol. 60, no. 1, p. 11, 2017. [Online].

Available at: https://doi.org/10.1145/3014349

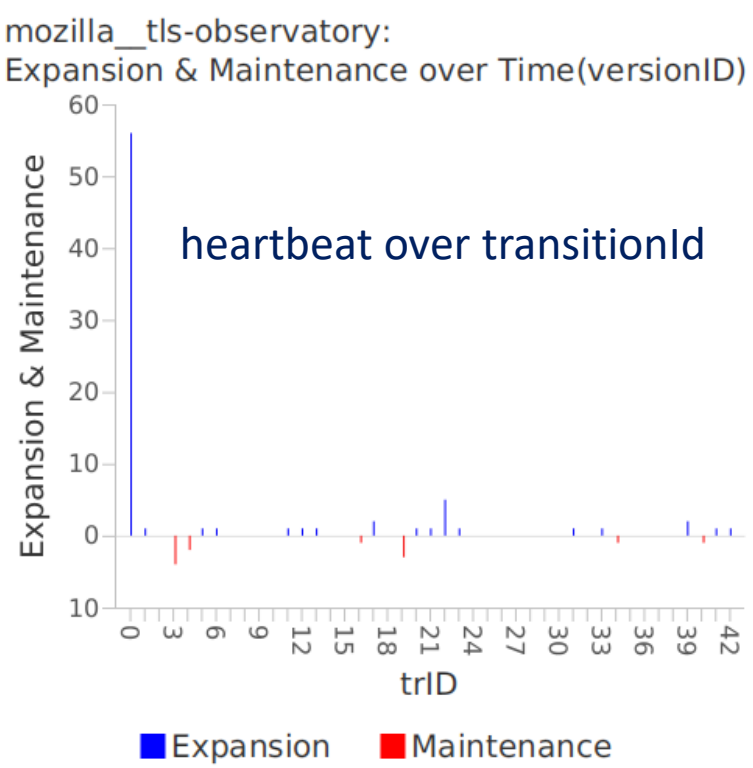
Stats for “Antarctica”

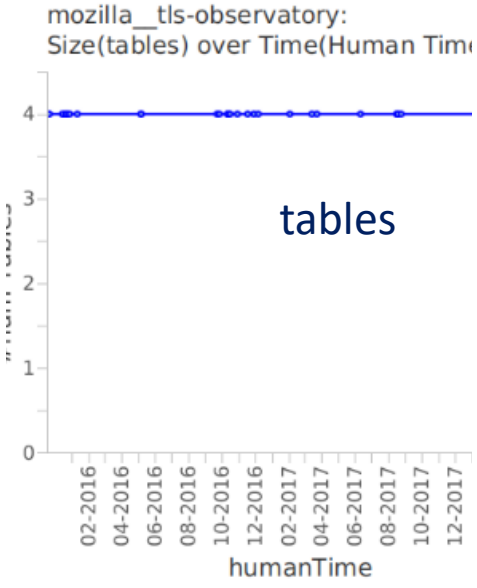
	Frozen				Almost Frozen				Fshot n Frozen			
Count	34				65				25			
	min	med	max	avg	min	med	max	avg	min	med	max	avg
Sch. Upd. Period (months)	1	1	69	8.24	1	6	99	11.98	1	2	46	9.28
TotalActivity	0	0	0	0	1	3	10	3.62	11	23	383	45.64
#Commits	2	2	11	3.18	2	3	13	3.83	2	4	17	4.56
#Active Commits	0	0	0	0	1	1	3	1.40	1	2	3	1.76
#Reeds	0	0	0	0	0	0	0	0	0	1	3	0.84
Turf commits	0	0	0	0	1	1	3	1.40	0	1	3	0.92

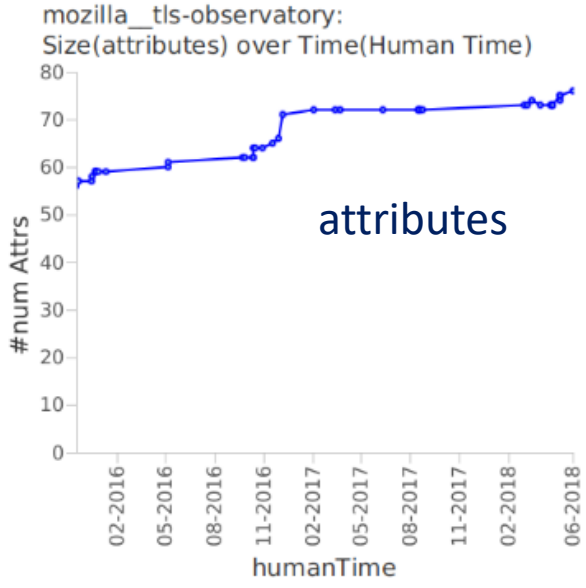
Moderate

29 projects, with the following **median** values:

- Schema Update Period: 20 months
- #Commits: 10, 7 of them active, typically all of them turf
- Total change: 23 attributes.
- Schema line: 65% of projects with a rise in the schema, 10% with a flat line, the rest: turbulent or dropping
- Proj. duration (months): 72% with > 24, 86% with > 12







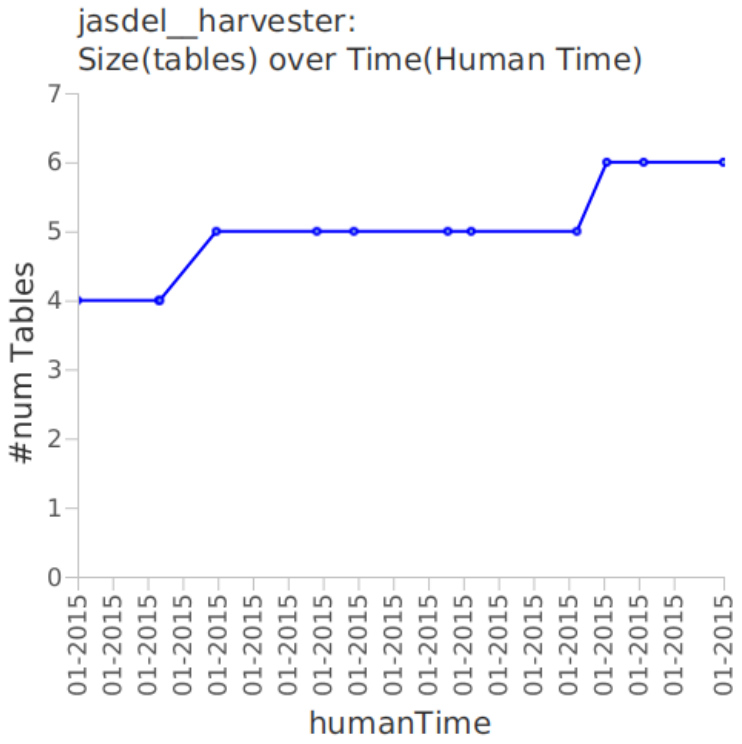
schema size over human time

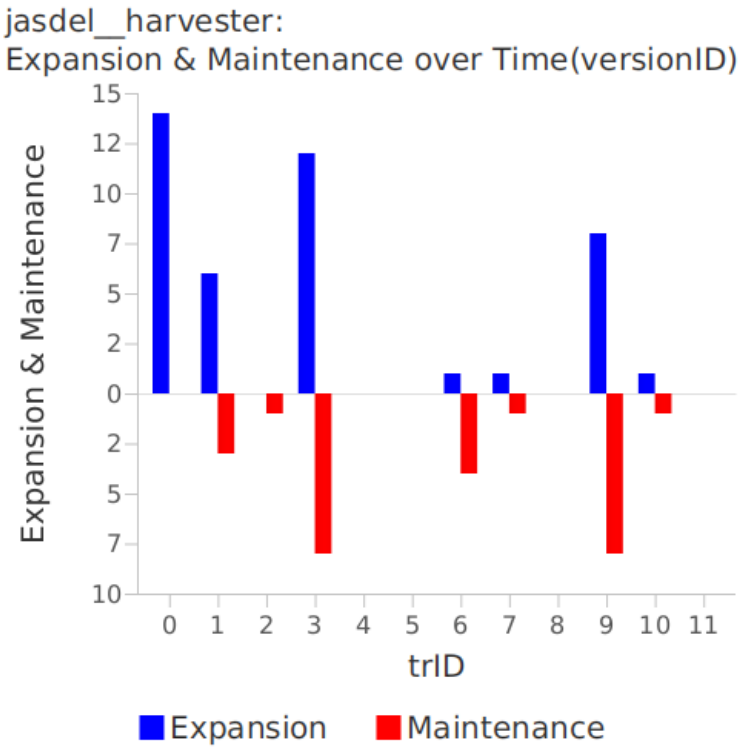
Focused Shot & Low

Median values for 20 projects:

- Schema Update Period: 17.5 months
- Commits: 10.5 commits, 6.5 of them active, with ~1 reed
- Total activity: 71 attributes (!!)
- Proj. duration (months): 70% of the projects > 24, 75% > 12 months.

Change mostly due to 1-2 " reeds" (vs. regular, small volume "turf" of moderate) can rise to significantly higher volumes than previous taxa





Active

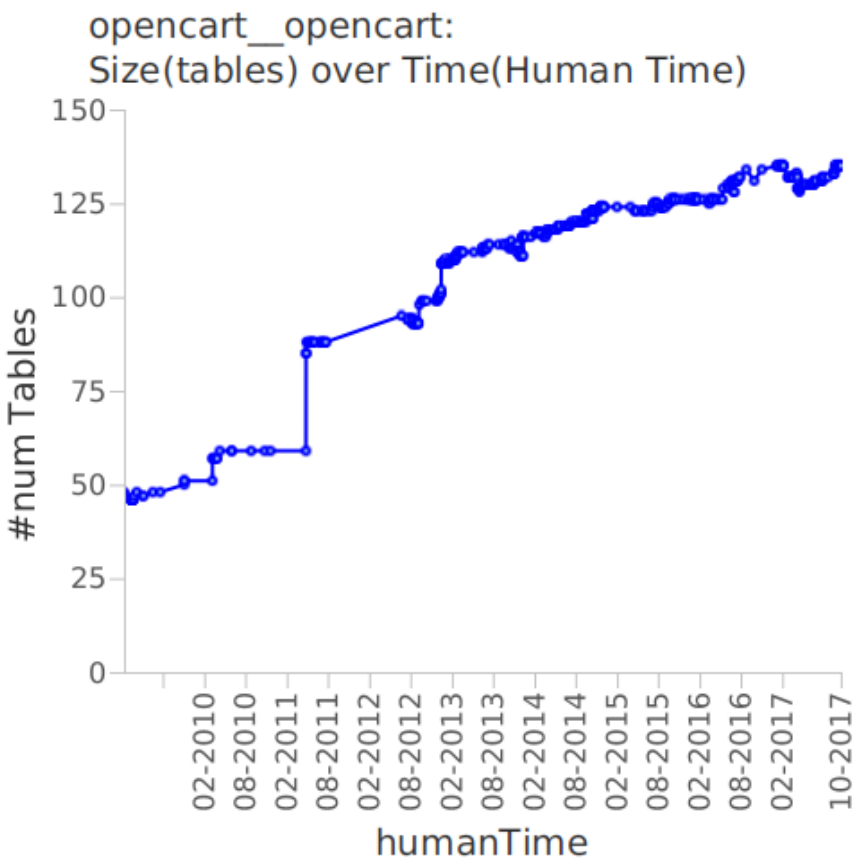
Median values for 22 schemata with **significant volume of updates**.

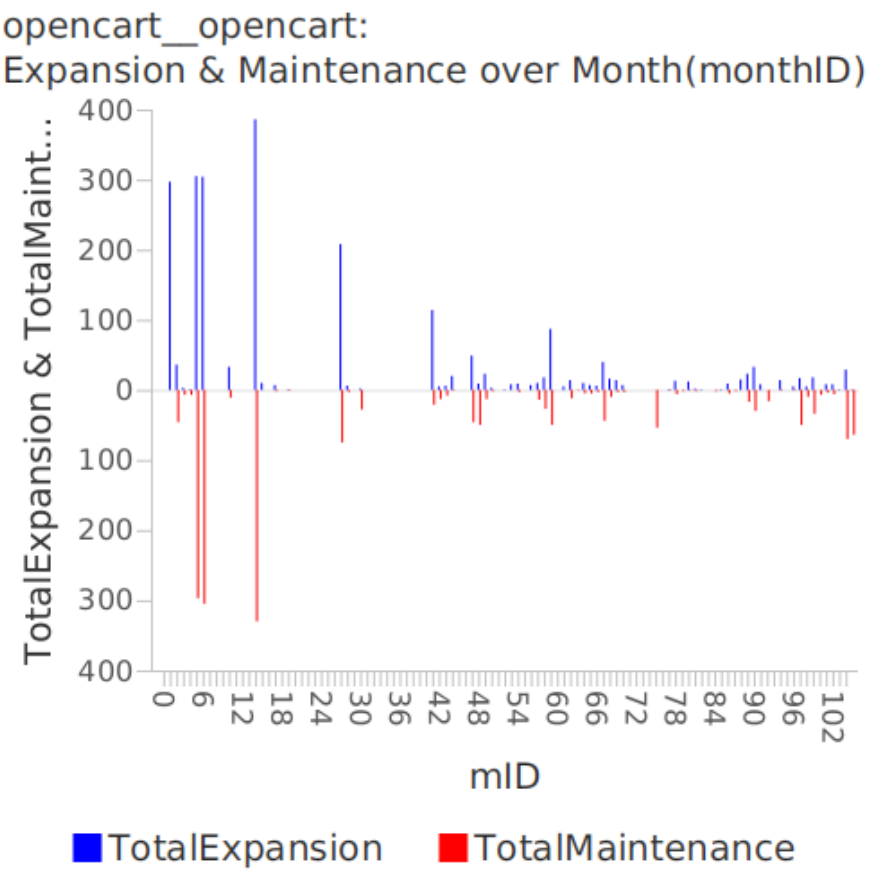
SUP durations: 31 months

Heartbeat: 36.5 commits, 22 active, 5.5 reeds and the rest turf,

Total activity: 254 attributes

Project duration (months): 91% of the proj. > 24, 95% > 12 months.

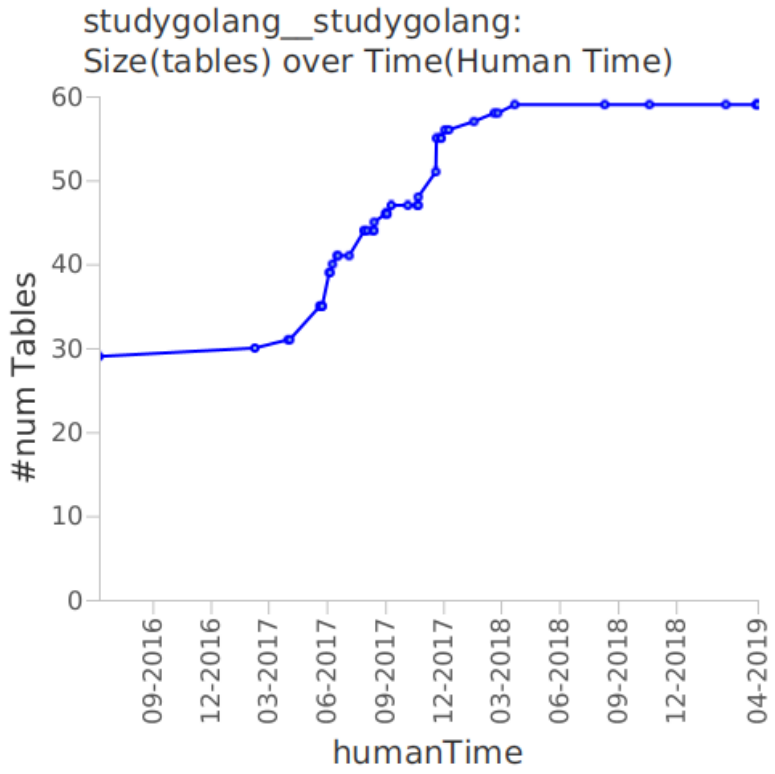


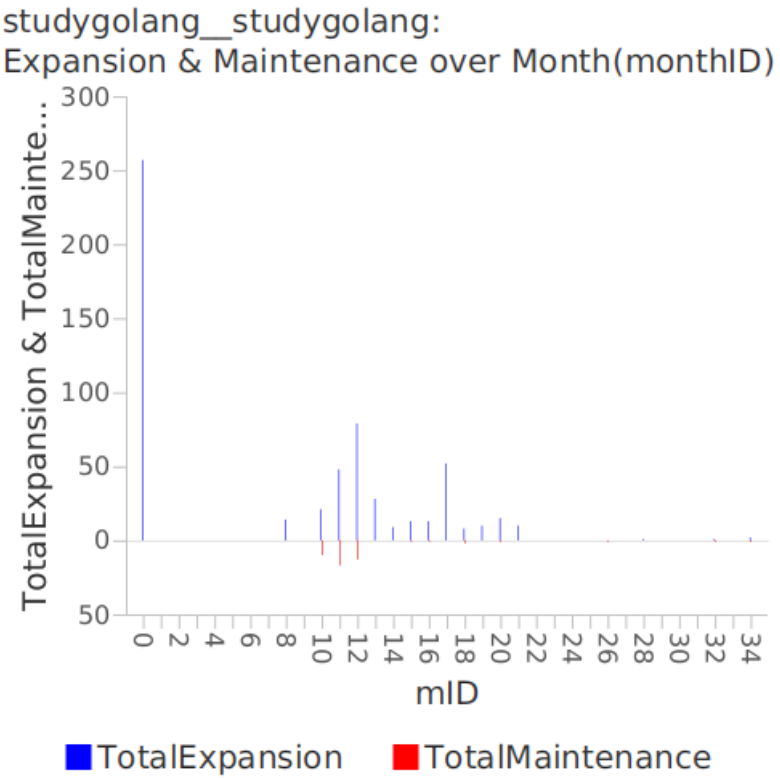


Active

The heartbeat of the 22 prj's is **not homogeneous**:

- **Frequency**: periods of systematic turf activity, periods of idleness, spikes of massive maintenance, growth and restructuring.
- **Schema size**: typically growing (50% of cases: multi-step, 9% with a single step); also 2 cases of flat schemata, 3 cases of massive drop and 4 cases of turbulent evolution of schema size.





Stats for moderate – active taxa

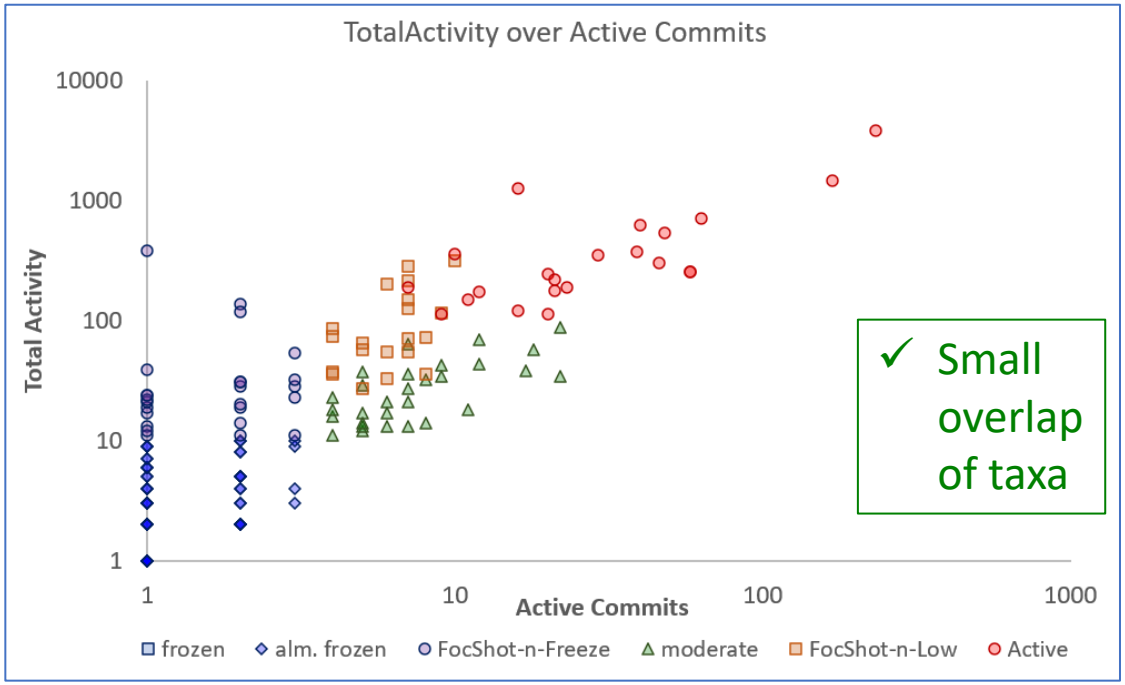
	Moderate				Fshot n Low				Active			
Count	29				20				22			
	min	**med**	max	avg	min	**med**	max	avg	min	**med**	max	avg
Sch. Upd. Period (months)	1	**20**	100	23.62	1	**17.5**	57	21.05	1	**31**	100	35.95
TotalActivity	11	**23**	88	30.0	27	**71**	315	105.15	112	**254**	3485	546.14
#Commits	5	**10**	43	13.52	7	**10.5**	19	11.55	9	**36.5**	516	77.36
#Active Commits	4	**7**	22	8.52	4	**6.5**	10	6.30	7	**22**	232	43.95
#Reeds	0	**0**	2	0.17	1	**1**	2	1.40	1	**5.5**	31	7.32
Turf commits	4	**7**	22	8.34	2	**5**	9	4.90	0	**18.5**	207	36.60

Taxa Validation

Are these taxa reasonable, based on the data?

Properties of a well-defined set of taxa

- **Completeness**:, i.e., covering all possible cases of activity behavior
  - ✓ Covered both by the data and by the classification scheme
- **Disjointness**: the characteristics of the different taxa are different, and each project can belong to exactly one taxon
  - ✓Also covered by pairwise mutual exclusion of the constraints



- **Internal Cohesion**: within a taxon, the behavior of its projects is similar
  - … more difficult to prove… ->

Cohesion test #1: compare taxa over the entire data set

- We assessed the statistical significance of the taxa differences over (i) their number of active commits and (ii) their total activity via the Kruskal-Wallis test
  - The null hypothesis of the test is that the different taxa have the same median and thus the reported p-value is a measure on the rejectability of the null hypothesis
  - Activity measurements: Kruskal-Wallis chi-squared = 178.22, df = 5, p-value < 2.2e-16
  - Active Commits: Kruskal-Wallis chi-squared = 175.27, df = 5, p-value < 2.2e-16.

Cohesion test #2: comparing taxa pairwise

- We compared the taxa pairwise via a Kruskal-Wallis test.
- See the p-values of the respective test: the lower left triangle refers to the **active commits** and the upper right triangle to the **total activity**.
- Assuming an acceptance threshold of 5%, the test reveals that the differences between taxa are significant, with the exception of two cases.

… wrt total activity

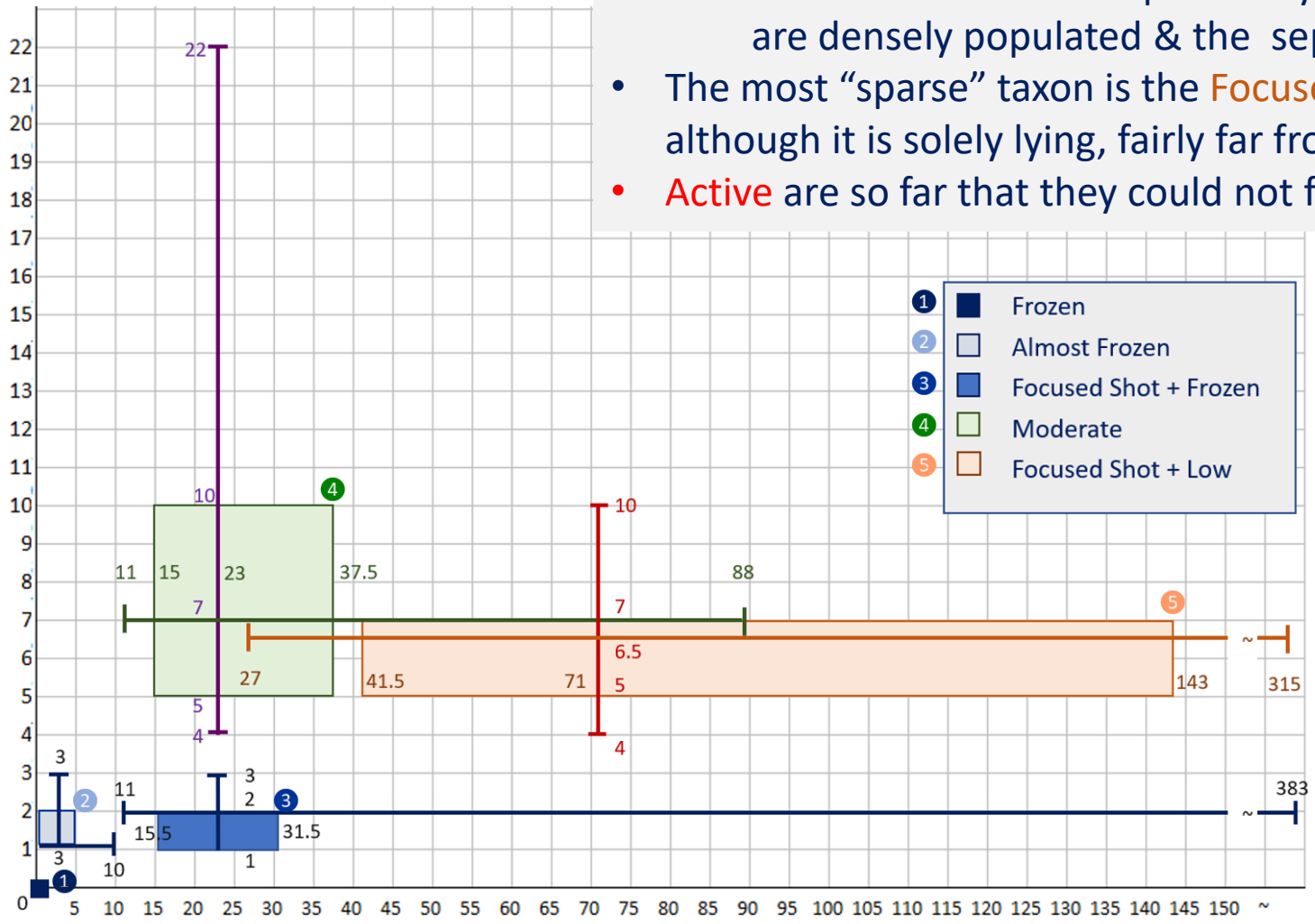
…wrt active commits

	Alm. Frozen	FShot+Frozen	Moderate	FShot+Low	Active
Alm. Frozen		1.730e-13	8.455e-15	1.141e-11	2.013e-12
FShot+Frozen	0.03199		0.7945	2.138e-05	6.076e-08
Moderate	3.714e-16	2.282e-10		5.406e-06	1.294e-09
FShot+Low	3.884e-13	7.043e-09	0.2796		1.855e-05
Active	7.204e-14	3.1103-09	5.355e-07	9.745e-08	

Cohesion test #3: visually depict the quartiles

We computed the quartiles of total activity and active commits for each taxon

- All taxa are heavily biased towards lower values, for both active commits and activity
- The 3 most frozen taxa are really clustered in very cohesive boxes (with the exception of few Focused Shot& Frozen).
  - The small surface is expected by definition, but they are densely populated & the separation is justified
- The most "sparse" taxon is the Focused Shot & Low, although it is solely lying, fairly far from the rest
- Active are so far that they could not fit in the image



Summary of Contributions

- A **clear definition of the nomenclature** and the **important measures** of the problem.
- The **compilation of a the largest (publicly available) dataset on schema evolution till now**
- **RQ1: Taxa do exist**
- **RQ2: Gravitation to rigidity: absence of evolution is more widespread than its presence**
- **RQ3: The frequency, volume and radical nature of change are low (with exceptions)**

RQ1. Is schema evolution present extensively?

- No! Specifically, **although evolution is present, its absence is way more omnipresent**
- Out of 327 identified projects with quality guarantees:
  - 40% had no schema evolution whatsoever …
  - an extra 10% had no change at the logical level …
  - an extra 20% were almost frozen

- This is in sharp contrast with the reported 50% - 70% of Software Maintenance effort
- *Conjecture: the absence of schema evolution is not due to the lack of its necessity, but rather due to its difficulty*

RQ2. Are there archetypal patterns of schema lives?

- Yes! Out of the 195 with at least a single commit:
  - Frozen projects (17%) with no change whatsoever
  - Almost Frozen (33%) with few active commits and small change
  - Focused Shot and Frozen (13%) with practically a single spike of change
  - Focused Shot and Low projects (10%) with a couple of high-volume reeds of evolution and less than 10 active commits overall.
  - Moderate projects (15%), with of constant rate of schema maintenance but less than 90 attributes changed in their lifetime
  - Active projects (11%) with frequent change and high volumes of it

RQ3. What are the demonstrable properties of schema evolution, in terms of volume, frequency and important characteristics?

- The clarification of nomenclature, units and measurement process is a contribution per se
- The measurement of evolution characteristics. With the exception of the active category (11% of the population)…
  - … **frequency** of change is really low in almost all taxa
  - … the **change in terms of tables** added (& esp.) deleted is small
  - … **focused massive updates** do exist, but are few
    - … in the order of 1 – 2 for non-frozen projects
    - … and in fact are present also in almost frozen projects

Open roads

- Elephant in the room: the absence of industrial schema histories (and thus our reliance on FOSS projects only)
- How to teach ~~our students~~ people better on how to design, evolve, and link-to-code relational schemata?
- If gravitation to rigidity is an inherent issue of the relational model & RDBMSs what can we do?
  - Progressively move to non-relational models?
  - Design better models of linking code to databases?
- In any case, continue research with respect to
  - … the why's of gravitation to rigidity
  - … the patterns of evolution, in the lives of both tables and schemata

Everything is online!!

https://github.com/DAINTINESS-Group/

My group's git page

https://github.com/DAINTINESS-Group/

has links to **Data sets**

https://github.com/DAINTINESS-Group/Schema_Evolution_Datasets/tree/master/SchemaEvolutionDatasets2020

and Code

… for computing differences (Hecate)

… visualizing schema lives (Plutarch Par. Lives)

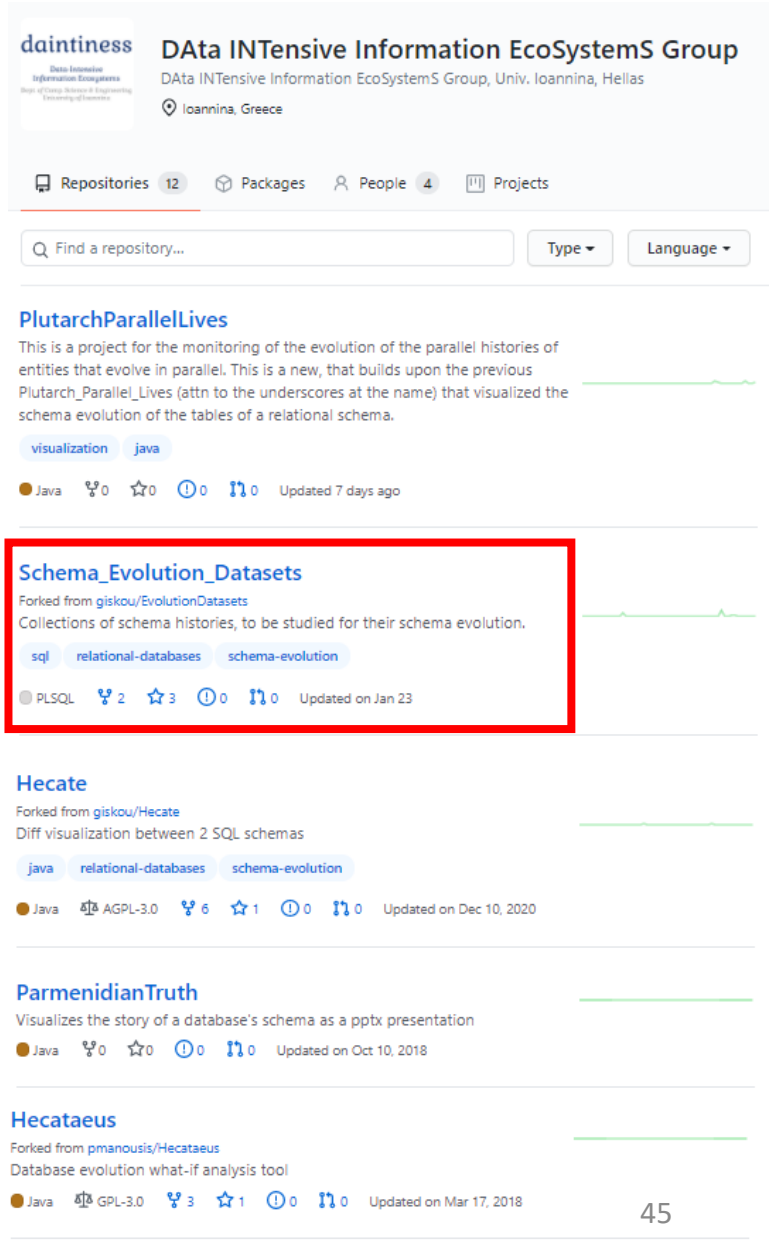
… visualizing the structure of FK's (Parmenidian Truth)

… handling the impact of evolution (Hecataeus)

My **Schema Biographies** web page

http://www.cs.uoi.gr/~pvassil/projects/schemaBiographies/

has links to **Papers & Results**



Thank you!

- Many thanks go to …
  - Apostolos Zarras, Petros Manousis,
  - Ioannis Skoulis, Fanis Giahos, Michael Kolozoff, Athanasios Pappas, Maria Zerva, Konstantinos Dimolikas, Theologia Kalakou,
  - Savvas Kostoudas, Alexandros Voulgaris, Nikolaos Pantelidis …
- … for all the work done on Schema Evolution, in the Univ. Ioannina, since 2013

To probe further (code, data, details, presentations, …)
www.cs.uoi.gr/~pvassil/projects/schemaBiographies/