

# Provenance Systems Identified

In this document we include an extract of the study performed as part of the systematic review presented in [8]. In particular, after performing a specific search strategy of papers related to the provenance context, we have identified a list of 251 provenance systems existing in the literature, together with the number of their appearances within the different papers found in the search.

The following tables 1, 2 and 3, present the set of provenance systems identified. More specifically, per each system we show its name (including its citation when no name is given to the system), together with the number of its appearances. It is important to note that the systems identified with numbers from 1 to 25 (see grey cells in Table 1) correspond to the provenance systems we have analysed in [8].

| Ref. | System          | Num. | Ref. | System                             | Num. | Ref. | System              | Num. | Ref. | System             | Num. |
|------|-----------------|------|------|------------------------------------|------|------|---------------------|------|------|--------------------|------|
| 1    | myGrid/ Taverna | 93   | 26   | PANDA                              | 11   | 51   | StarFlow            | 5    | 76   | Pin                | 3    |
| 2    | kepler          | 90   | 27   | ReproZip                           | 11   | 52   | Triana              | 5    | 77   | Pip                | 3    |
| 3    | VisTrails       | 88   | 28   | noWorkflow                         | 10   | 53   | Chiron              | 4    | 78   | Ghoshal et al. [1] | 3    |
| 4    | PASS            | 73   | 29   | Mondrian                           | 9    | 54   | Geo-Opera           | 4    | 79   | Sahoo et al. [2]   | 3    |
| 5    | Trio            | 41   | 30   | Git2prov                           | 8    | 55   | IWBrowser           | 4    | 80   | RDFProv            | 3    |
| 6    | Chimera         | 29   | 31   | ParaView                           | 8    | 56   | LabelFlow           | 4    | 81   | RunMyCode.org      | 3    |
| 7    | Karma           | 26   | 32   | SAM                                | 8    | 57   | Marathe             | 4    | 82   | SystemT            | 3    |
| 8    | Cui 2000        | 25   | 33   | Web Provenance                     | 8    | 58   | S-Language          | 4    | 83   | Tariq et al. [3]   | 3    |
| 9    | PASOA/PreServ   | 24   | 34   | CDE                                | 7    | 59   | TDB                 | 4    | 84   | TREC               | 3    |
| 10   | SPADE           | 21   | 35   | Haystack                           | 7    | 60   | YesWorkflow         | 4    | 85   | Tupelo             | 3    |
| 11   | ZOOM            | 21   | 36   | Janus                              | 7    | 61   | Arab et al.         | 4    | 86   | WDO-It!            | 3    |
| 12   | swift           | 18   | 37   | OPUS                               | 7    | 62   | ARK                 | 3    | 87   | ZOO                | 3    |
| 13   | Wings-Pegasus   | 18   | 38   | Provmanager                        | 7    | 63   | Artemis             | 3    | 88   | Gibson et al. [4]  | 2    |
| 14   | ORCHESTRA       | 17   | 39   | SOLE                               | 7    | 64   | Bochner et al. 2008 | 3    | 89   | ATOM               | 2    |
| 15   | Tioga           | 17   | 40   | Vansummeren et al. [5]             | 7    | 65   | DataHub             | 3    | 90   | Bonsai             | 2    |
| 16   | ES3             | 17   | 41   | General purpose provenance library | 6    | 66   | DataTracker         | 3    | 91   | Caravan            | 2    |
| 17   | Lipstick        | 15   | 42   | CMCS                               | 5    | 67   | DEEP                | 3    | 92   | Condor             | 2    |
| 18   | DBNotes         | 15   | 43   | Galaxy                             | 5    | 68   | eScience Central    | 3    | 93   | EEPS               | 2    |
| 19   | COMAD           | 14   | 44   | GenePattern                        | 5    | 69   | Flogger             | 3    | 94   | ExSpan             | 2    |
| 20   | Perm            | 14   | 45   | Ipython                            | 5    | 70   | Knitr               | 3    | 95   | Gaea               | 2    |
| 21   | Buneman [7]     | 12   | 46   | pFOD                               | 5    | 71   | libdft              | 3    | 96   | Geodise            | 2    |
| 22   | PLUS            | 12   | 47   | ProvBench                          | 5    | 72   | LogicBlox           | 3    | 97   | Geolineus          | 2    |
| 23   | RAMP            | 12   | 48   | Provenance Explorer                | 5    | 73   | Milieu              | 3    | 98   | GOOSE              | 2    |
| 24   | Redux           | 12   | 49   | SciCumulus                         | 5    | 74   | Oozie               | 3    | 99   | Hadoopprov         | 2    |
| 25   | Burrito         | 12   | 50   | SPIDER                             | 5    | 75   | ourSpaces           | 3    | 100  | IncPy              | 2    |

TABLE 1  
Provenance Systems and their appearances(l)

| Ref. | System                   | Num. | Ref. | System      | Num. | Ref. | System                | Num. | Ref. | System                      | Num. |
|------|--------------------------|------|------|-------------|------|------|-----------------------|------|------|-----------------------------|------|
| 101  | NetLogo                  | 2    | 126  | Aurora      | 1    | 151  | GExpLine              | 1    | 176  | Merlin                      | 1    |
| 102  | Oracle Total Recall (DB) | 2    | 127  | Azkaban     | 1    | 152  | Giovanni              | 1    | 177  | MetaML                      | 1    |
| 103  | PAPEL                    | 2    | 128  | BDBMS       | 1    | 153  | GridDB                | 1    | 178  | Windows Workflow Foundation | 1    |
| 104  | Prefuse                  | 2    | 129  | CAMERA      | 1    | 153  | GridSpace             | 1    | 179  | Minebench                   | 1    |
| 105  | ProTracer                | 2    | 130  | Caps        | 1    | 155  | gryPhyN               | 1    | 180  | Minemu                      | 1    |
| 106  | ProvAbs                  | 2    | 131  | CAPTRA      | 1    | 156  | Gspan and close-graph | 1    | 181  | ML                          | 1    |
| 107  | Provenance map orbiter   | 2    | 132  | CARE        | 1    | 157  | Hi-Fi                 | 1    | 182  | MMS                         | 1    |
| 108  | Provenancecurious2       | 2    | 133  | CARMEN      | 1    | 158  | Hive                  | 1    | 183  | MPO                         | 1    |
| 109  | ProvenanceJS             | 2    | 134  | Causeway    | 1    | 159  | Hyperloglog           | 1    | 184  | Myexperiment                | 1    |
| 110  | Story Book               | 2    | 135  | CCDB        | 1    | 160  | IADB                  | 1    | 185  | NetTrails                   | 1    |
| 111  | Subzero                  | 2    | 136  | CenterTrack | 1    | 161  | Ibis                  | 1    | 186  | Odyssey-SCM                 | 1    |
| 112  | Sweave                   | 2    | 137  | Century     | 1    | 162  | InsightNotes          | 1    | 187  | Opql                        | 1    |
| 113  | Travos                   | 2    | 138  | Clario      | 1    | 163  | IWBase                | 1    | 188  | OSLO                        | 1    |
| 114  | TrustCloud               | 2    | 139  | Collage     | 1    | 164  | JANUS                 | 1    | 189  | Panorama                    | 1    |
| 115  | UV-CDAT                  | 2    | 140  | CRISTAL     | 1    | 165  | Kairos                | 1    | 190  | Pastry                      | 1    |
| 116  | VIEW                     | 2    | 142  | Cytoscape   | 1    | 166  | Kieker                | 1    | 191  | PeerTrust                   | 1    |
| 117  | ViNE                     | 2    | 142  | D3S         | 1    | 167  | Knime                 | 1    | 192  | PiCloud                     | 1    |
| 118  | Vismashup                | 2    | 143  | Dexy        | 1    | 168  | LabBase               | 1    | 193  | PlanetFlow                  | 1    |
| 119  | W7                       | 2    | 144  | Dremel      | 1    | 169  | LBL                   | 1    | 194  | PMAF                        | 1    |
| 120  | Myers et al. [6]         | 1    | 145  | Dytan       | 1    | 170  | Log4j                 | 1    | 195  | POMELO                      | 1    |
| 121  | ACCESSPROV               | 1    | 146  | ESP2Net     | 1    | 171  | MACEDON               | 1    | 196  | pomsets                     | 1    |
| 122  | Amazonia                 | 1    | 147  | FilmTrust   | 1    | 172  | Mars                  | 1    | 197  | PRISM                       | 1    |
| 123  | Antfarm                  | 1    | 148  | Flashback   | 1    | 173  | Matrioshka            | 1    | 198  | Prizms                      | 1    |
| 124  | A-R-E                    | 1    | 149  | Flexiq      | 1    | 174  | MCA                   | 1    | 199  | PROB                        | 1    |
| 125  | Ariadne                  | 1    | 150  | GATES       | 1    | 175  | Mendel                | 1    | 200  | Propeller                   | 1    |

TABLE 2  
Provenance Systems and their appearances(II)

| Ref. | System             | Num. | Ref. | System     | Num. | Ref. | System      | Num. | Ref. | System    | Num. |
|------|--------------------|------|------|------------|------|------|-------------|------|------|-----------|------|
| 201  | ProPub             | 1    | 214  | ReTrace    | 1    | 227  | Sig.ma      | 1    | 240  | TIUPAM    | 1    |
| 202  | Provenance Browser | 1    | 215  | ReVirt     | 1    | 228  | Sourcetrac  | 1    | 241  | Tracefs   | 1    |
| 203  | ProvGen            | 1    | 216  | S2Logger   | 1    | 229  | SPROV       | 1    | 242  | Tribeca   | 1    |
| 204  | PROVglish          | 1    | 217  | SAF        | 1    | 230  | Spyglass    | 1    | 243  | TrustMe   | 1    |
| 205  | ProvStore          | 1    | 218  | SAGA       | 1    | 231  | StatJR      | 1    | 244  | Unicorn   | 1    |
| 206  | ProvToolbox        | 1    | 219  | SciPhy     | 1    | 232  | Subdue      | 1    | 245  | Utopia    | 1    |
| 207  | PubFlow            | 1    | 220  | Scorpion   | 1    | 233  | SUSE        | 1    | 246  | Valgrind  | 1    |
| 208  | PUG                | 1    | 221  | SecureFlow | 1    | 234  | Tabulator   | 1    | 247  | Vle-wfbus | 1    |
| 209  | Qfix               | 1    | 222  | seL4       | 1    | 235  | Talos       | 1    | 248  | W RBAC    | 1    |
| 210  | Quill++            | 1    | 223  | SElinks    | 1    | 236  | TAP         | 1    | 249  | WASA      | 1    |
| 211  | Rasdaman           | 1    | 224  | Sequoia    | 1    | 237  | TelegraphCQ | 1    | 250  | WiDS      | 1    |
| 212  | RDataTracker       | 1    | 225  | SHARE      | 1    | 238  | Tiresias    | 1    | 251  | Ymaldb    | 1    |
| 213  | RecProv            | 1    | 226  | Sieve      | 1    | 239  | Titian      | 1    |      |           |      |

TABLE 3  
Provenance Systems and their appearances(III)

## References

- [1] D. Ghoshal and B. Plale, "Provenance from log files: A bigdata problem," in Proceedings of the Joint EDBT/ICDT 2013 Workshops, ser. EDBT '13. ACM, 2013, pp. 290–297.
- [2] S. S. Sahoo, R. S. Barga, J. Goldstein, and A. P. Sheth, "Provenance algebra and materialized view-based provenance management," in Proceedings of the 2rid International Provenance and Annotation Workshop. Berlin: Springer, 2008, pp. 531–540.
- [3] D. Tariq, M. Ali, and A. Gehani, "Towards automated collection of application-level data provenance," in Proceedings of the 4<sup>th</sup> Workshop on the Theory and Practice of Provenance (TaPP'12). USENIX Association, 2012.
- [4] A. Gibson, M. Gamble, K. Wolstencroft, T. Oinn, C. A. Goble, K. Belhajjame, and P. Missier, "The data playground: An intuitive workflow specification environment," Future Generation Comp. Syst., vol. 25, no. 4, pp. 453–459, 2009.

- [5] S. Vansummeren and J. Cheney, "Recording provenance for SQL queries and updates," *IEEE Data Eng. Bull.*, vol. 30, no. 4, pp. 29–37, 2007.
- [6] J. D. Myers, L. Marini, P. Bajcsy, R. Kooper, Y. Liu, R. McGrath, J. Futrelle, T. McLaren, A. Collier, and A. Rodriguez, "A Digital Synthesis Framework for Virtual Observatories," *AGU Fall Meeting Abstracts*, Dec. 2008.
- [7] P. Buneman, A. Chapman, and J. Cheney, "Provenance management in curated databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Chicago, Illinois, USA, June 27-29, 2006, 2006, pp. 539–550.
- [8] B. Pérez, C. Sáenz-Adán, J. Rubio, A systematic review of provenance systems, submitted for publication.