Sharing statistics for SPARQL Federation optimization, with emphasis on benchmark quality

Kjetil Kjernsmo <kjekje@ifi.uio.no>

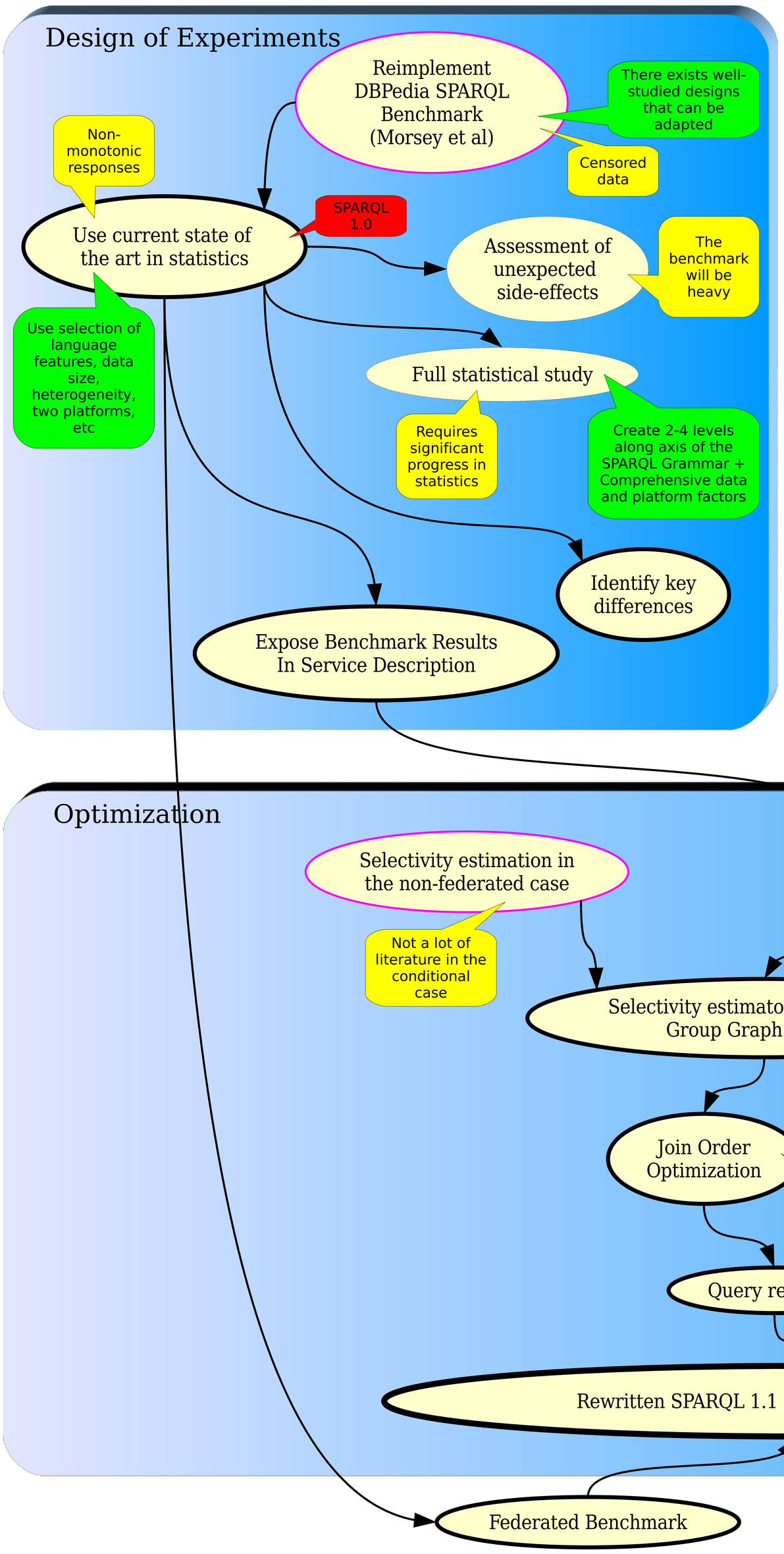
Department of Informatics, University of Oslo, Norway

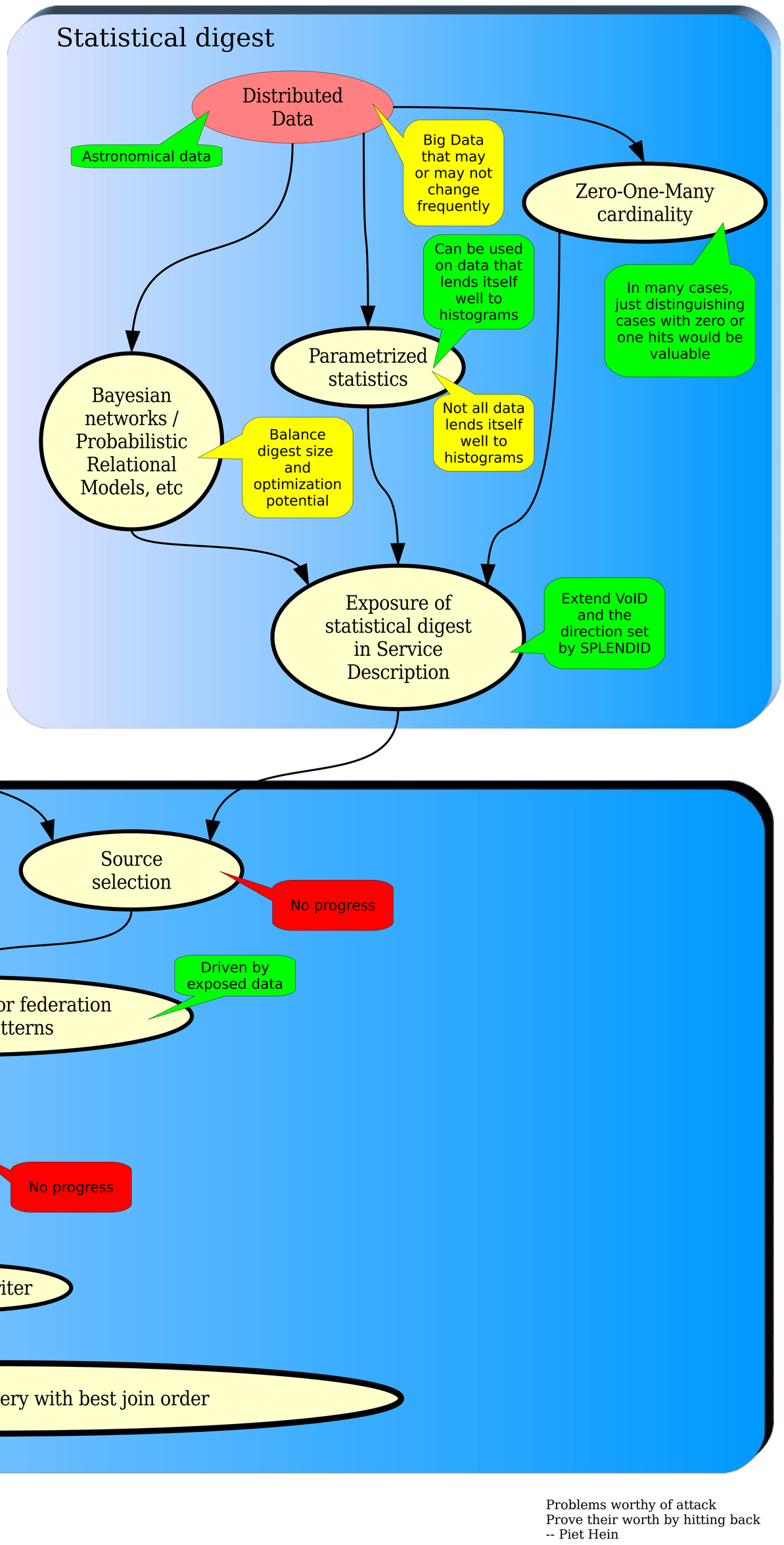
Key Problems:

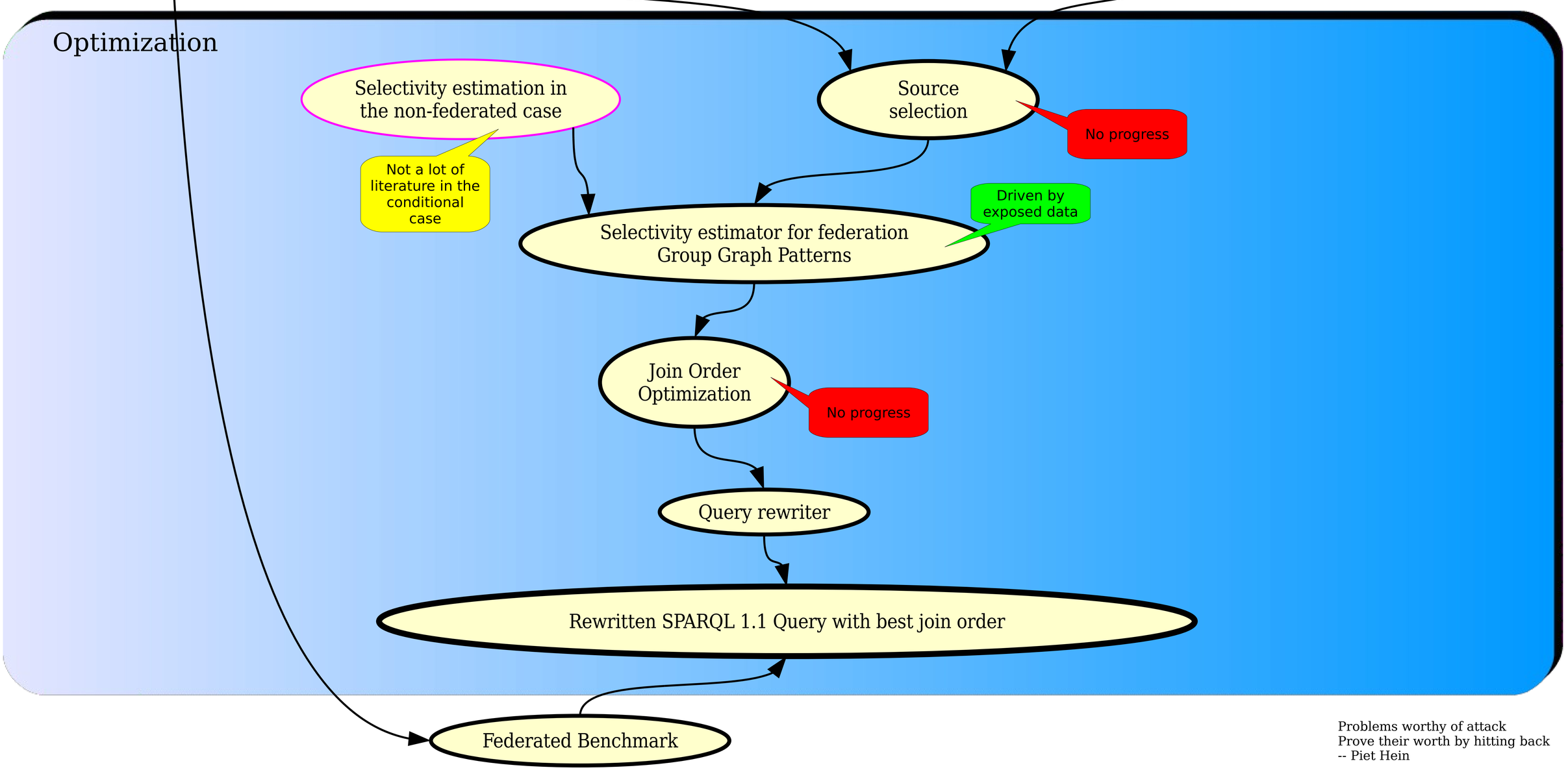
- Current benchmarks cannot meaningfully summarize different micro- or macrobenchmarks, thereby making falsification difficult.
- Advising users of SPARQL Endpoints on which implementation to use is difficult.
- Better statistics is needed for efficient join-order optimization in the federated case.
- Selectivity estimation for SPARQL needs to be improved.

Expected contributions:

- Benchmarks that are able to cover all realistic performance-influencing parameters for SPARQL engines and make it possible to weigh different parameters, as well as quantify unexplained differences.
- To assist optimization, enable endpoint service descriptions to expose:
  - performance characteristics,
  - statistical digests of data that are optimized with respect to size and query performance.
- SPARQL engine developers can use the benchmark to reliably quantify unexpected adverse effects from a given modification.
- New SPARQL engine users can identify key differences between different engines, and therefore be able to more wisely choose engine to use.







Problems worthy of attack
Prove their worth by hitting back
-- Piet Hein