Introduction

- Do you believe these outrageous headlines? Maybe they’re true? Wouldn’t it be useful to be able to check both the sources and calculations that lead up to their publication?
- Data deluges us — each day 2.5 quintillion bytes of data (2.5 · 10¹⁸) are created and eventually incorporated, either directly or indirectly, into other data items.
- Our options for checking their sources and accuracy are primitive. We depend on data, but cannot easily verify if it is dependable.
- In FRESCO we are building systems for automating data provenance capture, management and query, providing insights into the sources and computations that have produced a given data item.

Asking the right questions

- Where does this information come from?
- Under what conditions is it valid?
- How was it interpreted, transformed and processed?
- Does it depend on other pieces of data?

All of these are valid questions that allow us to reason about claims made using the data, whether they are published in a newspaper article or in a peer-reviewed journal.

Conclusion

Provenance is still a fledgling field of research in Computer Science. We believe the work of the FRESCO project will result in general purpose systems that allow users to more confidently reason about the sources and accuracy of their data.

Our ultimate goal is to make provenance support a fundamental aspect of all general purpose computing systems. FRESCO has wider implications beyond the project. Support for data provenance as a first class construct of modern computing systems will enable users to reason more confidently about the origin of data and verify the correctness of the computations carried out to derive it. This, in turn, should allow users to categorise data according to trustworthiness and ultimately result in an increase in the quality of data on which computations, and by extension, decisions are based.