Trust issues have been always dominant in the world of the Internet, no one believes everything that is out there, but rather replies on context, authority and context. If the data source cannot be directly trused

If the underlying data cannot be directly trusted, then the user should be suspicious of the answer(s) derived from them. Returning the source of individual pieces of information that lead to the answer(s) allows the user to verify the trustworthiness of the data. This is how the Web has always worked... no one believes everything they read on the Web, but rather relies on provenance and context.

The unique problem for Linked Data warehouses is that they consider the data as forming One Big Graph (or so is often the goal). As such, borders of provenance can become blurred, especially if inference is applied across these datasets. Different sources of data will be of varying quality on the Web, and if not careful, inference can smear crap from the untrustworthy sources all over the high-quality data. Thus, modularity of provenance is important. If inferencing is applied over domain A and domain B, then this should be tracked (unfortunately, not quite as easy as it sounds).

Furthermore, a lot of link-based analyses (like PageRank) can come to the rescue when taking heterogeneous, mixedquality Linked Data. Google has always been quite successful at finding "high-quality" sources of data for users using such techniques. Linked data engines can also use such techniques, assuming sufficient linking mechanisms are in place for the raw data.

Also, I would love to see Linked Data vocabularies making more use of all that juicy OWL stuff for pinpointing nonsense. I would love to see vocabularies defining irreflexive properties, and disjoint classes, and disjoint properties, and asymmetric properties, and functional datatype properties, and all that other goodness that can be used to automatically find inconsistencies and problems in the data, which can then be automatically fixed or flagged for users when used to derive answers. Such features are really not used enough at the moment. (In fact, perhaps OWL is not enough, and other mechanisms for implementing [local] quality/consistency checks are needed.)

The greater linked data community needs to develop vocabulary terms for expressing metrics for data quality -- consider the ten points above -- and then within individual communities develop agreed-upon means to determine those values. Arguably this is the "Dublin Core" approach to the problem, in the sense that terms like *completeness* or *consistency*would be reused across domains with inherently different domain-specific meanings, but such reuse would facilitate consumers from other communities choosing datasets. "The physics community says this dataset is accurate, by their measure.

Finally, getting users on board to rate and tag and edit information is an excellent way to collaboratively curate data. (Slight problem with this approach: finding users.)