

Data Usage Vocabulary Status Notes

March 13, 2015 status report

Eric Stephan

Bernadette Farias Lóscio

Illustrative Use Cases and Requirements from Use Case Document

It should be possible to cite data on the Web

- [UK Open Research Forum](#)
 - *Fits the scientific publication paradigm nicely feedback and source citing*
- [LA Times' Reporting of Ron Galperin's Infographic](#)
 - *source citing*

It should be possible to track the usage of data

- [ASO: Airborne Snow Observatory](#)
 - *Interesting because of the needs from an application perspective.*
- [UC Land Portal](#)
 - *Internationalization qualities for feedback*

Data consumers should have a way of sharing feedback and rating data.

- [RDESC](#)
 - *Need for collaborative feedback and rating data based on value to consumers*
 - *Has aspect of advertising data*
- [Open Experimental Field Studies](#)
 - *Need to collaborative feedback from wide community.*

Proposed Related Requirements (aren't part of data usage but will be used by Data Usage Vocabulary)

From Data Quality:

R-QualityOpinions

Subjective quality opinions on the data should be supported

R-QualityMetrics

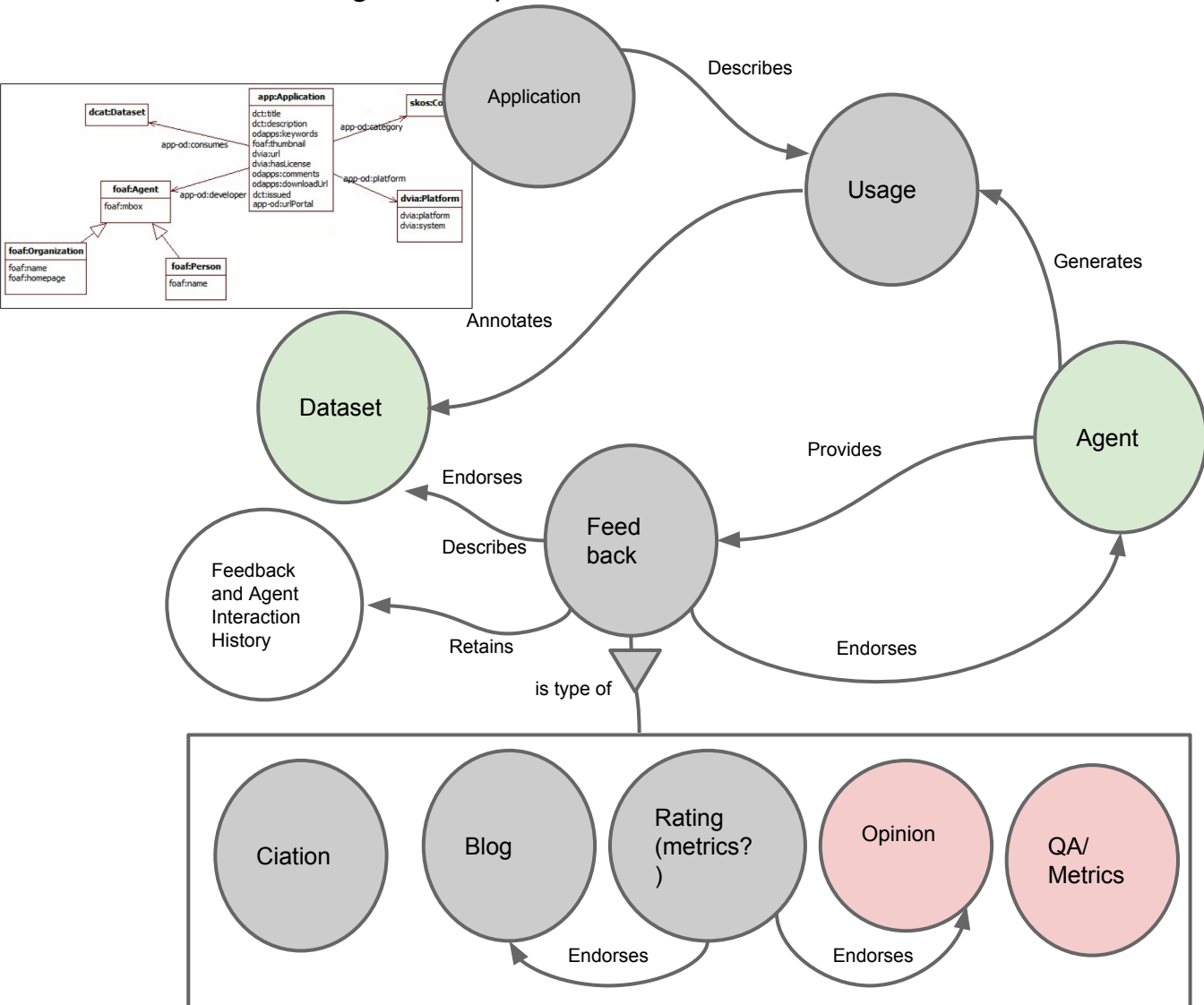
Data should be associated with a set of documented, objective and, if available, standardized quality metrics. This set of quality metrics may include user-defined or domain-specific metrics.

From Data Granularity:

R-GranularityLevels

Data available at different levels of granularity should be accessible and modelled in a common way

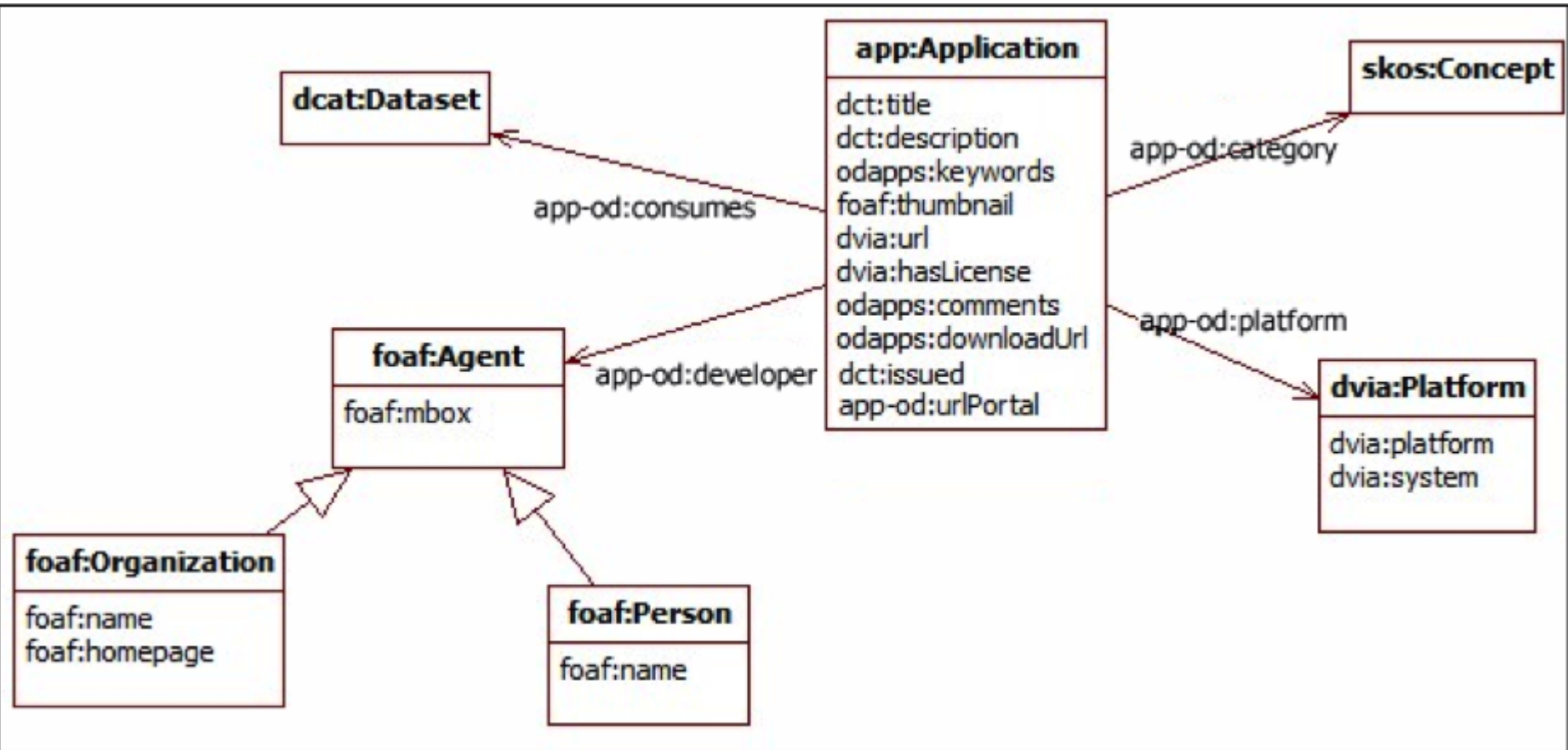
Data Usage Conceptual Model and Ties to Other Models



Data Usage Vocab
Data Quality -future
DCAT
PROV-O

Agent Interactions
 publisher <--> consumer
 producer <--> consumer
 consumer <--> consumer

Application Vocabulary (from Bernadette)



Proposed Data Usage Associations

**** denotes use of other vocabulary**

- Publisher is a type of ****AGENT**
- Producer is a type of ****AGENT**
- Consumer is a type of ****AGENT**
- ****AGENT** describes **USAGE**
- ****AGENT** provides **FEEDBACK**
- ****Application Vocabulary** describes **USAGE**
- **USAGE** annotates **DATASET**
- **FEEDBACK** endorses ****AGENT**
- **FEEDBACK** describes ****DATASET**
- **FEEDBACK** endorses ****DATASET**
- **FEEDBACK** retains ****PROVENANCE** (Feedback and Agent Interaction History)
- ****CITATION** is a type of **FEEDBACK**
- **BLOG** is a type of **FEEDBACK**
- ****OPINION** is a type of **FEEDBACK**
- ****QA and Metrics** is a type of **FEEDBACK**
- **RATING** is a type of **FEEDBACK**
- **RATING** endorses **BLOG**
- **RATING** endorses ****OPINION**

Observations about Data Usage Model

- Its a “glue” model assembling multiple existing (and soon to be existing) concepts
 - Data Quality
 - DCAT
 - Citations (there are many ways to cite) what do we suggest?
 - Provenance
 - Application vocabulary
- Data usage is focused on
 - humans / organizations (Agents) annotating datasets
 - Agents endorsing each other’s annotations
- How do we handle alternative definitions to dataset?:
 - e.g. file or API as DCAT created by a single agent
 - climate simulation results consist of thousands of files all considered a “data set”.