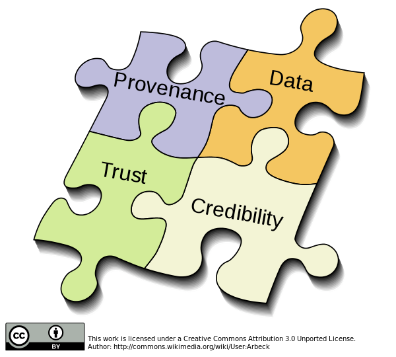
Provenance

Context:

ERA OF CONNECTIVITY – PEOPLE AND DEVICES

HOME COOKING VS STREET FOOD – what you know vs what you don’t (Good or bad?)

* Have to make decisions on data we don’t trust

How to make informed decision:

* Accountable:
  + “required or expected to

justify actions or decisions”

* Transparent and accountable systems
  + increase a participant’s understanding of the machine’s processes, and
  + can increase the participant’s trust of the system
* A description of what participants do
  + enables participants to feel that their actions can be monitored
  + incentive to conduct themselves in a respectable manner.

Data provenance – description of the way the data is collected and maintained by builders of algorithms, accompanied by and exploration of potential bias. Public scrutiny provides maximum opportunity for correction, but privacy must be respected

World Wide Web Consortium Provenance Definition - Provenance is a record that describes the people, institutions, entities, and activities, involved in producing, influencing, or delivering a piece of data or a thing in the world

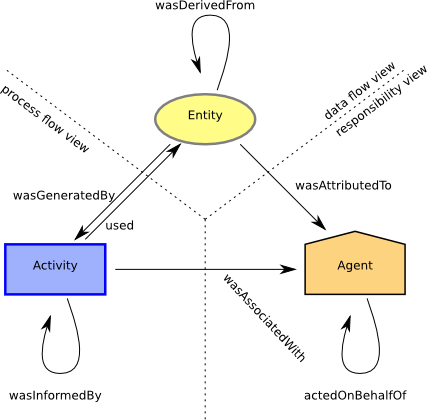
STANDARD OF PROVENANCE = PROV

Provenance working group – sharing of provenance between datasets

* PROV: standardised vocabulary to describe
  + Flow of data
  + Processes
  + Responsibility
* PROV allows systems to be enriched with “data lineage” showing the origin of data
* A huge step towards systems accountability

THREE CORE CONCEPTS:

|  |  |
| --- | --- |
| **Social Machine** | **PROV** |
| Piece of information, decision, vote, document | ENTITY |
| Actions such as voting, writing, reporting, commenting, approving | ACTIVITY |
| Person, service, system, organization, collective | AGENT |



* Provenance may provide too much information so summaries are needed to understand provenance. Requirements:
  + **Essence** of Provenance: a provenance summary should capture the essence of the provenance graph that it summarises.
  + **Outliers**: It should be possible to detect anomalies or outliers in a provenance summary
  + **Conformance**: It should be possible to decide whether a provenance graph is compatible, or conformant, with a provenance summary.
* Another problem with provenance is trusting the data, you have to get data from participants that have an interest in the subject and that are going to provide reliable data