



Enabling Assessment and Propagation of Information Quality in the South Esk Hydrological Sensor Web

Heiko Müller, Qing Liu, Quan Bai, Corne Kloppers, Andrew Terhorst
Research Scientist
07th April 2011

South Esk River Catchment



The South Esk

- 214 km long (longest river in Tasmania).
- Catchment area ~3,350 km².
- Rainfall varies between 500mm-1,800mm.
- Water flows into a hydro-electric storage dam.

South Esk Hydrological Sensor Web



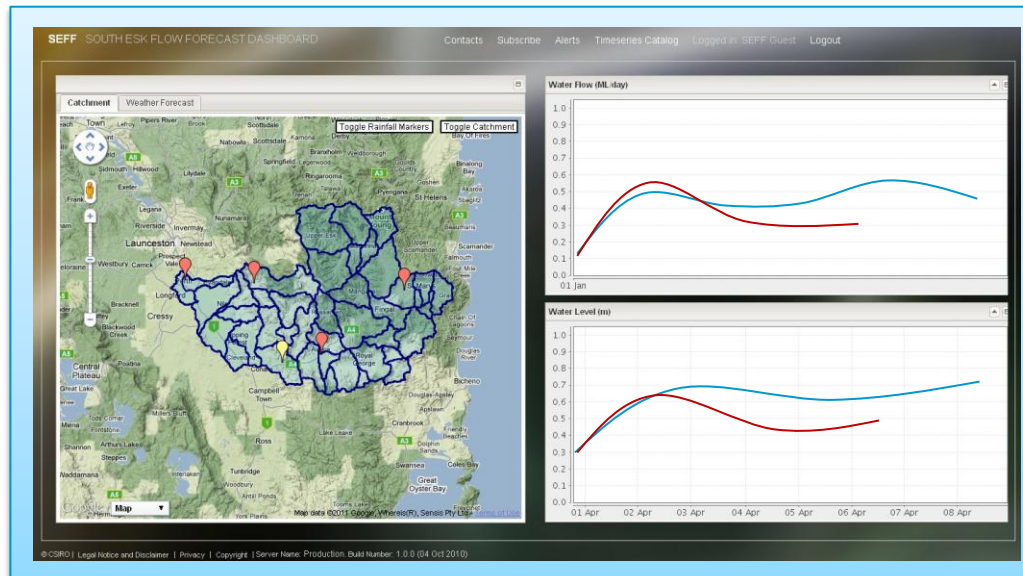
Sensing in the South Esk

- Over 30 different sensors.
 - Rainfall, Wind speed, Air temperature, ...
- Owned by different agencies
 - CSIRO, BOM, Hydro Tasmania, DPIPW.
- Integration based on OGC's Sensor Web Enablement (SWE).

South Esk Hydrological Sensor Web

- Observations from the aggregated sensor assets drive a rainfall-runoff model that predicts river flows at key monitoring points.
- Maintain environmental flows and sustain a hydro-power facility at the bottom of the catchment.
 - Apply Water Restrictions.
 - Announce Flood Take (Temporarily Lift Water Restrictions).

South Esk Flow Forecast



South Esk Flow Forecast

- Runs every 2 hours.
- Estimates water flow and water level at four monitoring points.
- Generates automated alerts.



Information Quality

Quality information in the South Esk Sensor Web

Manage uncertainty and assess the quality of information is essential to assist in decision making and help establish trust.

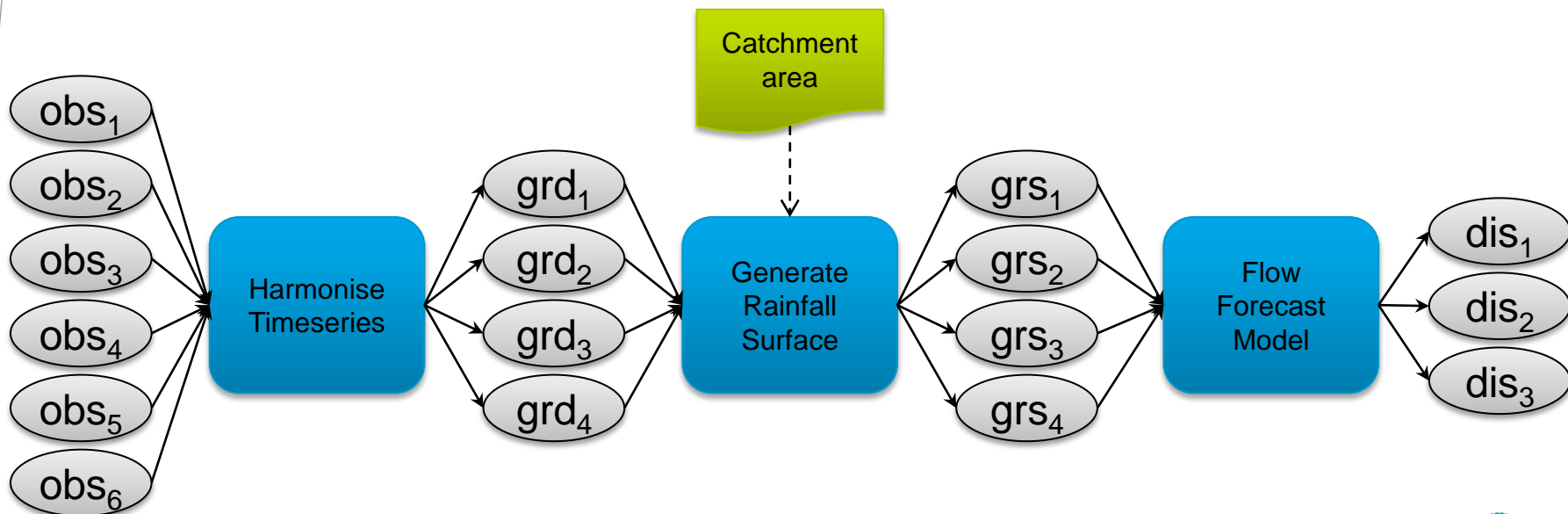
- Four areas of research in data curation can help in dealing with information quality and uncertainty in environmental systems.
 - **Citation:** Ensure that every data item has a unique identifier.
 - **Provenance:** Metadata about the process that generated a piece of information.
 - **Versioning:** Maintain the history of changes to a database.
 - **Annotation:** Attach annotations to individual data items and propagate them to derived data values.

Data Provenance

Provenance of a resource is a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource.

Provenance XG Final Report, 2010.

Provenance is a graph that models data and process dependencies (and process execution context).

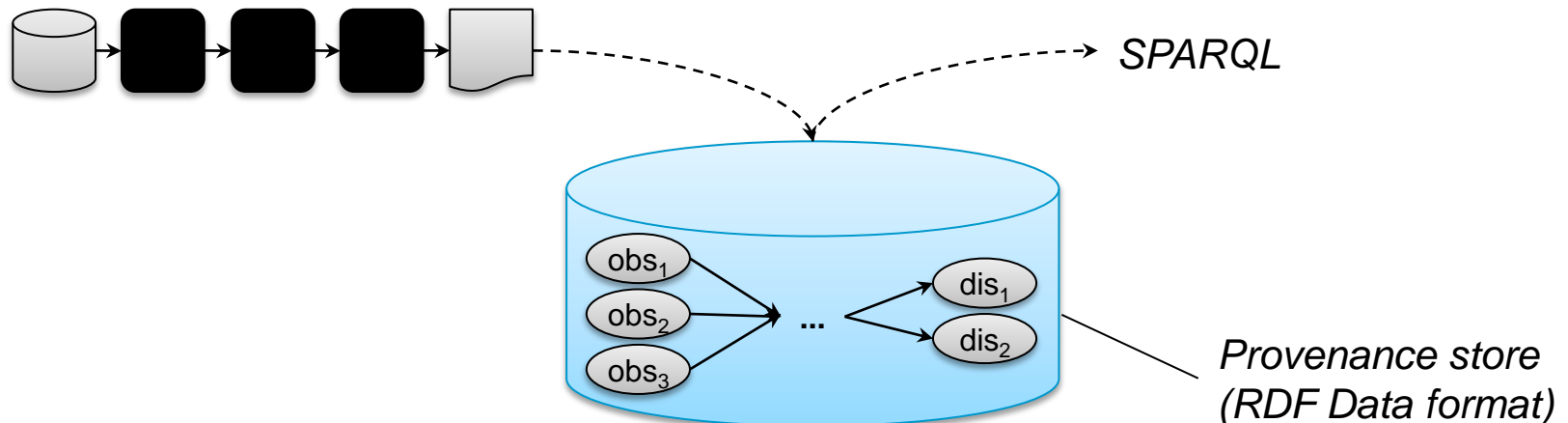


Data Provenance (cont.)

Provenance facilitates data quality assessment

- Make the data derivation process transparent to the user.
 - *Which model was used?*
 - *How many different sensors where used?*

Provenance management system

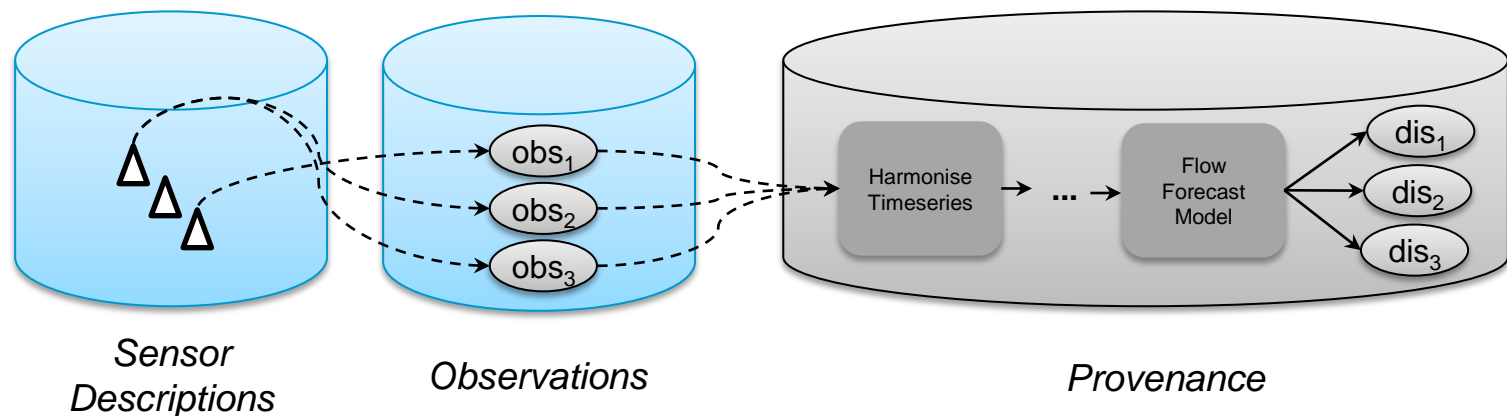


Versioning & Citation

Enable persistent data links

- Essential for linking data and to ensure that links remain stable under data evolution.
- Enable retrieval of additional metadata
 - Sensor type, location, last calibration date.
 - Data quality codes assigned by data providers.

Linked data store



Versioning & Citation (cont.)

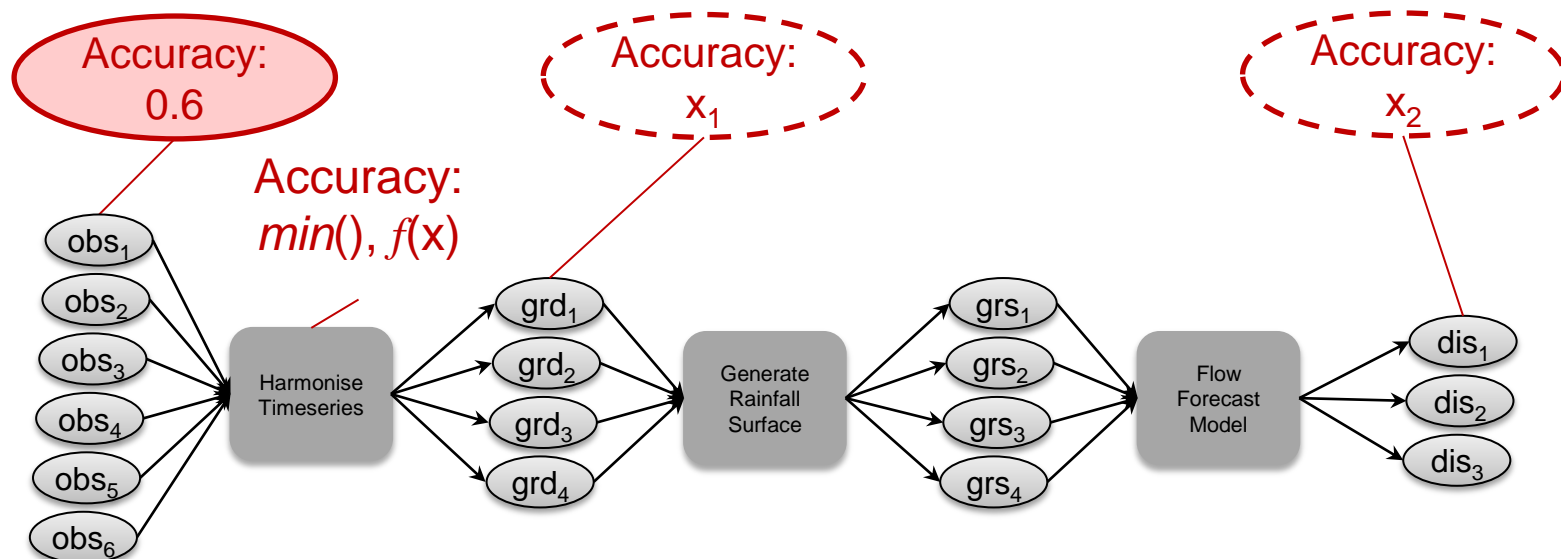
History of data is provenance information

- Observations and sensor descriptions are updated.
- Log who changed what and when.
 - Find all forecasts based on observations that changed since.

Annotation

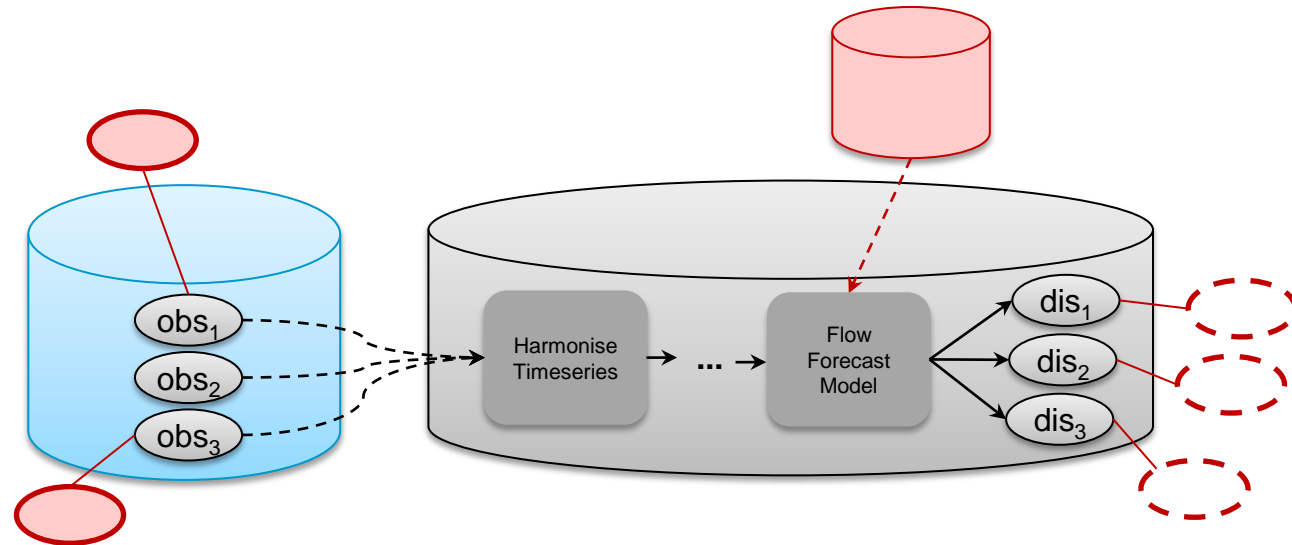
Maintain and propagate additional metadata

- Annotate individual data items using controlled vocabularies or domain ontologies.
 - Simple labels or key-value pairs.
- Define rule for propagating annotation for derived values based on provenance information.



Summary & Outlook

Assess and propagate information quality



Future work

- Machine learning to identify patterns in provenance information that indicate low-quality forecasts.
- Assess uncertainty from change history.
- Constraints on provenance and annotations to raise alerts.

Tasmanian ICT Centre

Heiko Müller
Research Scientist

Phone: +61 3 6232 5575

Email: heiko.mueller@csiro.au

Web: www.csiro.au/science/TasICTCentre.html

www.csiro.au

Thank you

Contact Us

Phone: 1300 363 400 or +61 3 9545 2176

Email: enquiries@csiro.au Web: www.csiro.au

