Why EML Metrics

- Primary quality checks are limited
    - schema compliance
    - EML parser (ids and references)
- Dataset quality not sufficient for automated use
    - EcoTrends
    - Site experiences

Data Manager Library

- Read and Parse EML datasets
- Download data entities, store in RDBMS
- Query with SQL-like constructs

Can be used to create the next level of quality control checks for EML datasets

Any valid EML can still be contributed, but limits of usability will be clear
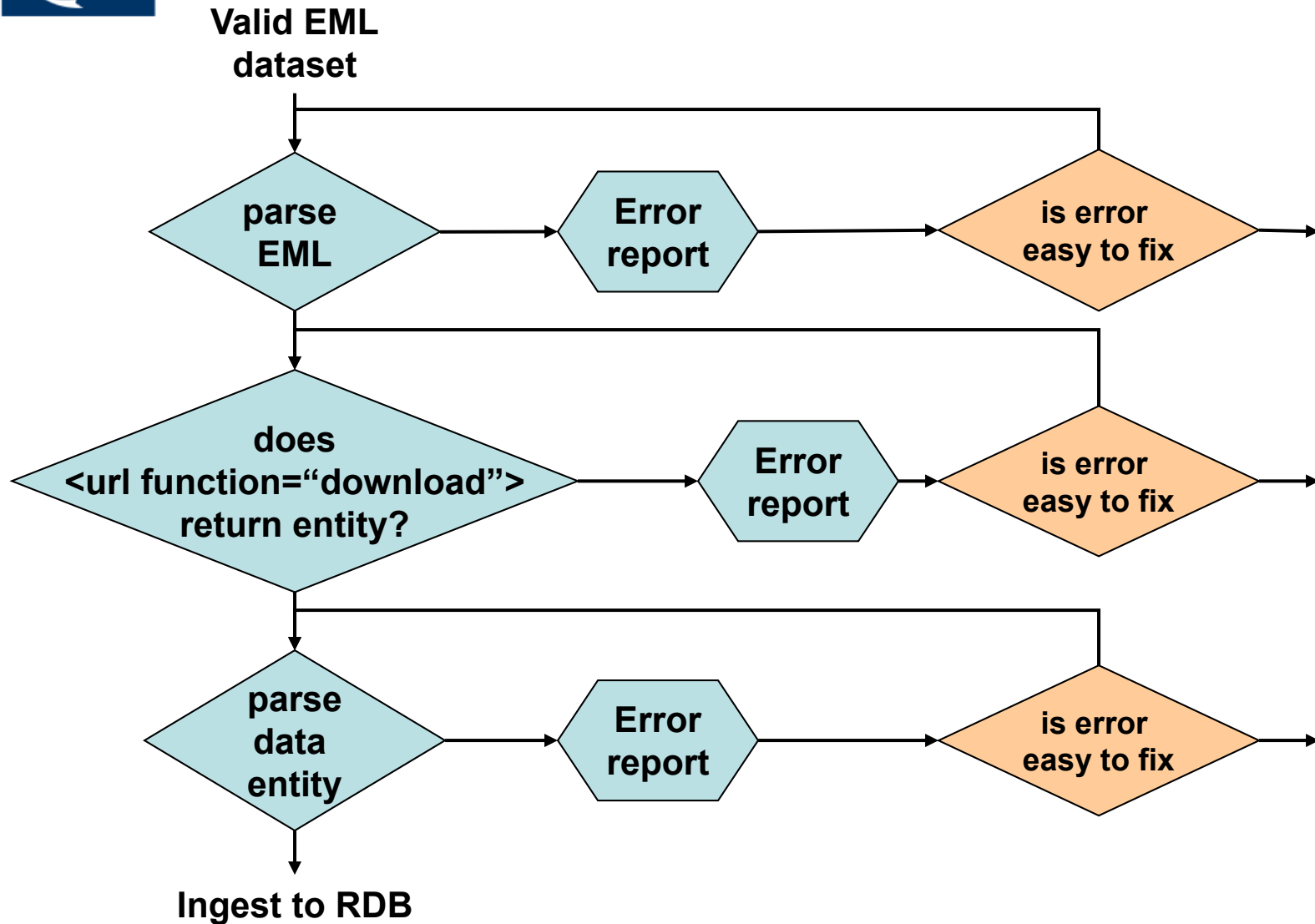
Reports

- Multiple levels of complexity
  - Success of data entity ingestion (y/n)
  - Describe entity (#cols & rows, delimiter, typing)
  - Data-metadata comparison

- LTER specific Reports
  - 2004 best practices recommendations
  - Use of network keywords, units, …

As a tool for site IMs

Valid EML dataset

parse EML → **Error report** → **is error easy to fix** →

does <url function="download"> return entity? → **Error report** → **is error easy to fix** →

parse data entity → **Error report** → **is error easy to fix** →

Ingest to RDB

Needed improvements to site system

Goals

- LTER EML Metrics group develops list of criteria (with IMexec)

- Software tools initiated (Duane)

- Initial reports on sites' datasets in LTER Metacat produced by September 2010 (IMC annual meeting)

Today's timeline and tasks

1000 – 1015: Introduction

1015 – noon: List metrics for EML data packages

A. features required for any EML data table to be read and ingested by the data manager library

B. features specific to LTER (e.g., compliance with Best Practices)

1300-1400: Software and initial report format

– Organizing feedback for Duane

– Report to IMC in September

1400-1500: Wrap up

– Identify needs to move forward (eg,IMC WG?, Tiger team?)

– Writing assignments or VTC schedule