COCOMO II: Analysis of SDSS Software Project

S. Kent, Fermilab

- What is COCOMO II?
- SDSS Development Costs
- Analysis using COCOMO II
- Caveats

What is COCOMO II?

- "COnstructive COst MOdel"
  - Model for estimating software cost, schedule, and effort
  - Developed at USC starting 1981
  - Based on source "lines of code"
  - Set of equations with large number of "tuning parameters"
    - Output is "person-months" of effort (excluding vacation)
  - Latest version is calibrated with 161 projects
  - Several software packages available to enter data and compute results
    - Windows, SunOS binaries freely available

What is covered?

- **COCOMO II includes:**
  - **Plans and Requirements**
  - **Product design**
  - **Programming**
    - **Detailed Design**
    - **Code and Unit Test**
  - **Test planning, verification and validation**
  - **Project Office**
  - **Code management/QA**
  - **Manuals**
  - **Maintenance**

What is not covered?

- COCOMO II does not include:
  - Planning and requirements as part of a larger project
  - Operations
    - System Administration
    - Production system
      - Hardware procurement
      - Software/Hardware deployment
      - Database creation
    - Data ingest
  - Creation of static datasets (e.g., organizing input astrometric catalogs)
  - Large simulation efforts
  - Probably other things

COCOMO Planning Phases

- **Early design model**
  - **System software/architecture not yet selected**
  - **Fewer tuning parameters**
- **Post-archicture model**
  - **System architecture selected**
  - **Mode of operation and cost risks understood**
  - **Full set of tuning parameters**

COCOMO Metrics

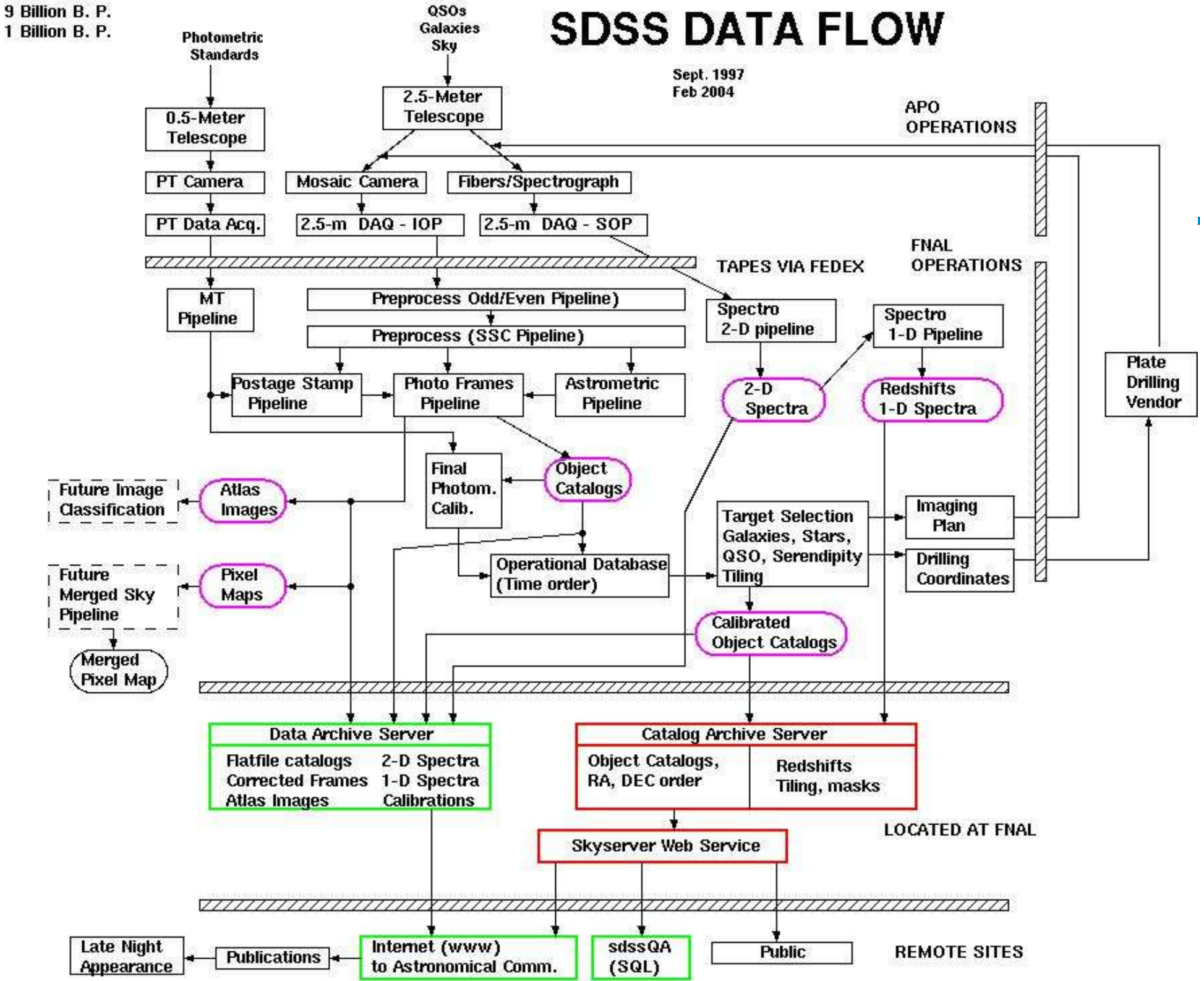
- **Three options for basic unit of measure:**
  - **Source line of code (Logical vs. Physical Lines)**
  - **Function points (roughly, 1 FP = 100 LOC)**
  - **Adaptation of existing code**
    - **Tuning parameters to measure, e.g., what fraction of code needs to be modified**
- **Code is organized in modules**
  - **Some tuning parameters are specific to a module**
  - **Others are common to a project**

Tuning Parameters

- **Scale factors (accounts for exponential dependence of project on LOC):**
  - **Precedentedness (has this ever been done before)**
  - **Team cohesion**
  - **...**
- **Effort multipliers**
  - **Reliability of code**
  - **Degree of documentation**
  - **Experience of programmers**
  - **Storage, execution time constraints**
  - **...**

SDSS Software Development

- Development cycle 1991-1999
- Included Activities
  - Data acquisition system
  - Data processing system
    - Imaging
    - Spectroscopy
    - Photometric Telescope
    - Production framework
  - First-generation data distribution system
  - All hardware (including custom VME boards)
- 160 FTE-years
- $16 million


9 Billion B. P.
1 Billion B. P.
Photometric Standards
QSOs Galaxies Sky
SDSS DATA FLOW
Sept. 1997
Feb 2004
0.5-Meter Telescope
2.5-Meter Telescope
APO OPERATIONS
PT Camera
Mosaic Camera
Fibers/Spectrograph
PT Data Acq.
2.5-m DAQ - IOP
2.5-m DAQ - SOP
TAPES VIA FEDEX
FNAL OPERATIONS
MT Pipeline
Preprocess Odd/Even Pipeline)
Preprocess (SSC Pipeline)
Spectro 2-D pipeline
Spectro 1-D Pipeline
Plate Drilling Vendor
Postage Stamp Pipeline
Photo Frames Pipeline
Astrometric Pipeline
2-D Spectra
Redshifts 1-D Spectra
Final Photom. Calib.
Object Catalogs
Future Image Classification
Atlas Images
Target Selection Galaxies, Stars, QSO, Serendipity Tiling
Imaging Plan
Drilling Coordinates
Operational Database (Time order)
Future Merged Sky Pipeline
Pixel Maps
Calibrated Object Catalogs
Merged Pixel Map
Data Archive Server
Flatfile catalogs
Corrected Frames
Atlas Images
2-D Spectra
1-D Spectra
Calibrations
Catalog Archive Server
Object Catalogs, RA, DEC order
Redshifts
Tiling, masks
LOCATED AT FNAL
Skyserver Web Service
Late Night Appearance
Publications
Internet (www) to Astronomical Comm.
sdssQA (SQL)
Public
REMOTE SITES

COCOMO II Applied to SDSS

- 34 New Modules, e.g. ...
  - Photometric pipeline
  - Target selection code
  - Astroda - data acqusition system
- 19 Modules adapted from preexisting code, e.g.
  - TCL
  - Objectivity
  - (List is incomplete)
- 1.3 million physical lines of source code equivalent
  - C, C++, TCL, SQL, ASP, JS

Tuning COCOMO's Parameters

- **Caveat - I use physical LOC, not logical, but exclude header files.**
- **COCOMO provides default tuning parameters. To reproduce SDSS results, I generally tune as follows:**
  - **1. Project has low reliability, documentation needs**
  - **2. Programmers are all experts with high programmer continuity**
  - **3. Development platform is stable, and interactions between geographically separate sites are good.**

Prediction

- **Total Lines of code equivalent: 1,341,714**
- **Person-months: Range of 1383-2161. Most likely is 1729 (157 FTE years, assuming 1 month vacation per year)**
  - **==> We are a factor ~7-10 more efficient than predicted by the nominal COCOMO model (!!!)**
- **$12 million for personnel (assumes $60K per year base salary plus 28% fringe)**
  - **==> $4 million for hardware, M&S, travel**
  - **(My rule of thumb is that HW, M&S, etc are 25% of personnel costs, which would predict $3 million)**
- **Above costs exclude overhead.**

Caveats

- **SDSS estimates are not entirely accurate**
  - **In-kind scientist time - some included, some not**
  - **LOC count not what COCOMO uses**
  - **Code count refers to today's codebase, not that in 1999.**
  - **Unclear if all activities are captured by COCOMO - e.g., code configuration management and deployment to remote sites.**
  - **Distinction between development, operations, and maintenance not entirely clear, and SDSS did all with essentially the same people.**