ENCODE
Encyclopedia of DNA Elements

ENCODE Applicant Information Meeting

December 18, 2006

Fishers Lane Conference Center

Rockville, Maryland

Peter Good, NHGRI

Goodp@mail.nih.gov

General Guidance

- Read the RFA very carefully and completely
- Formulate ideas and questions
- Contact Program Staff as early on as possible
  - Send ideas and questions by email first
  - Do not need to wait until Letter of Intent Deadline
- In proposal, address elements defined in the Research Scope

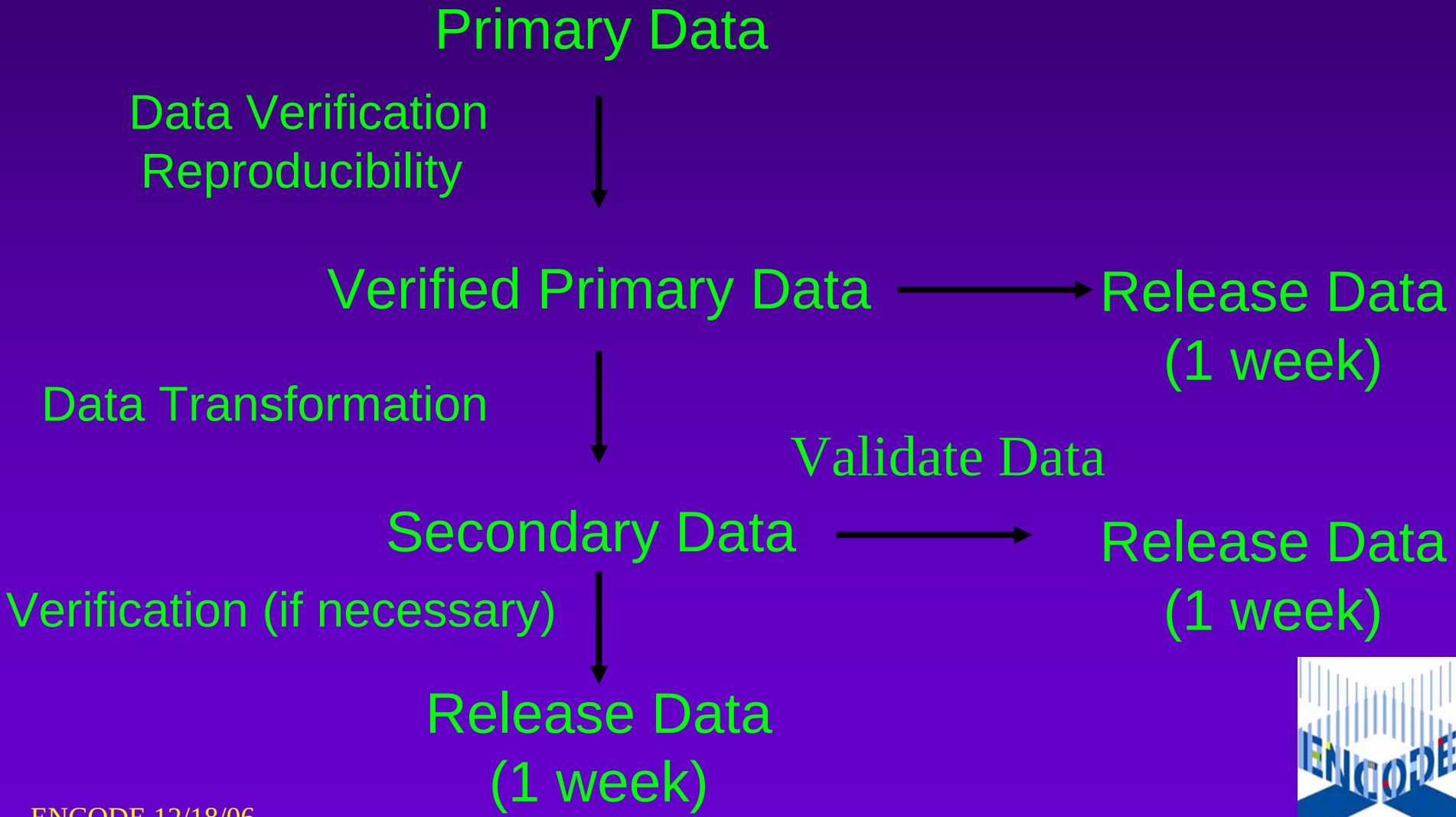
Next Phase of ENCODE

- Support efforts to apply high-throughput methods to develop a comprehensive catalog of functional elements in the human genome sequence (RFA HG 07-030)
- Support a data coordination center (DCC) to house and maintain the ENCODE data (RFA HG 07-031)
- Support similar efforts in selected model organisms: data production and DCC (modENCODE) (RFAs closed – funding March 2007)
- Continue to support technology development efforts (RFAs closed – funding June 2007)

Data for ENCODE

- Definitions for data release policy
  - Verification: Is the data reproducible?
    - Platform-specific standard
  - Validation: Is the data accurate?
    - A second assay to determine if the biochemical event is real
    - Performed on a fraction of the verified elements to determine quality of the dataset

Data Release Pipeline



Data for ENCODE DCC

- Primary data
  - Submitted to public databases
    - GEO, ArrayExpress, GenBank
  - DCC needs to track
- Secondary data (processed)
  - Elements extracted from primary data
    - Hit list of targets from ChIP-chip, etc
- Validation data
- Metadata: Information about the experiment
- Related data from public databases

DCC requirements

- Track data as it is produced
  - Reports to NHGRI staff
- Collect and store data from production centers
  - Efficient mechanisms to collect data in established formats
  - Robust data management tools
  - Worked with production centers on data exchange mechanisms
  - Must provide links back to primary data

DCC requirements

- Disseminate data
  - Multiple mechanisms
    - Biologist / Single gene users - Browser view
    - Power users - bulk downloads
  - Work with other informatics resources to disseminate data
- Participate as an active member in the ENCODE Consortium
  - Generate defined data freezes for analysis

DCC Infrastructure

- Robust data management tools
- Reuse of existing software where possible
- Must have acceptable resource sharing plan compatible with the ENCODE project

Data Release

- Must provide unencumbered access to the data produced by the ENCODE Consortium
- Data release policy evaluated by review group

General Guidance

- Read the RFA very carefully and completely
- Formulate ideas and questions
- Contact Program Staff as early on as possible
  - Send ideas and questions by email first
  - Do not need to wait until Letter of Intent Deadline
- In proposal, address elements defined in the Research Scope