

# **Session V: Life Science Identifiers - Use Cases, Future Directions**

# Recent History

- LSIDs 3 years old
- I3C evaluating AGAVE, BSML
  - encoded IDs as tuples/triples
- If we could not agree on a data standard, could we at least agree on how we write the identifiers

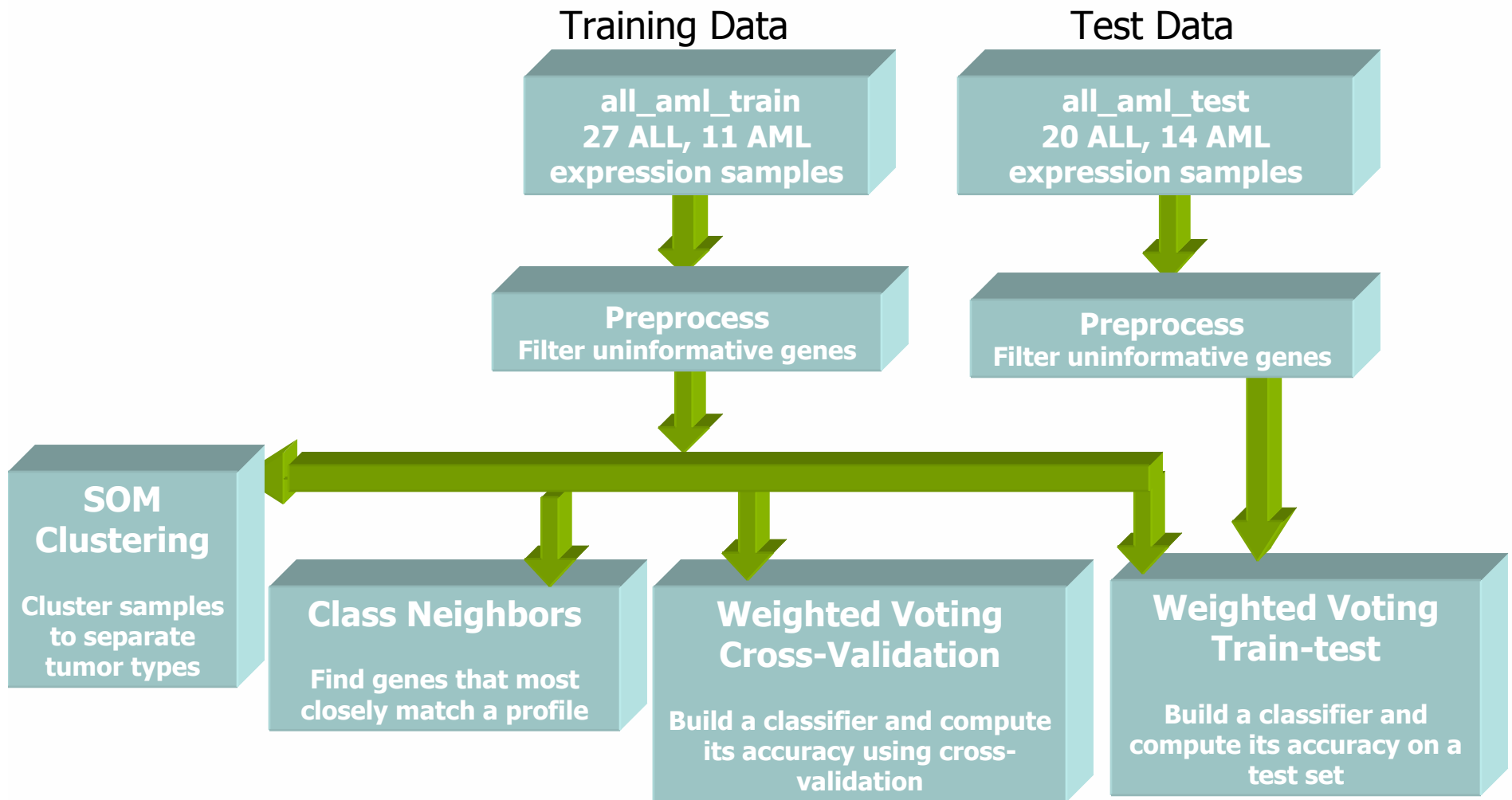
# Today

- OMG Spec
- google “+LSID +bioinformatics”
  - 686 results (10/27/04, 2:40pm)
  - 700 results (10/27/04, 7:20am)

# Broad Use Cases

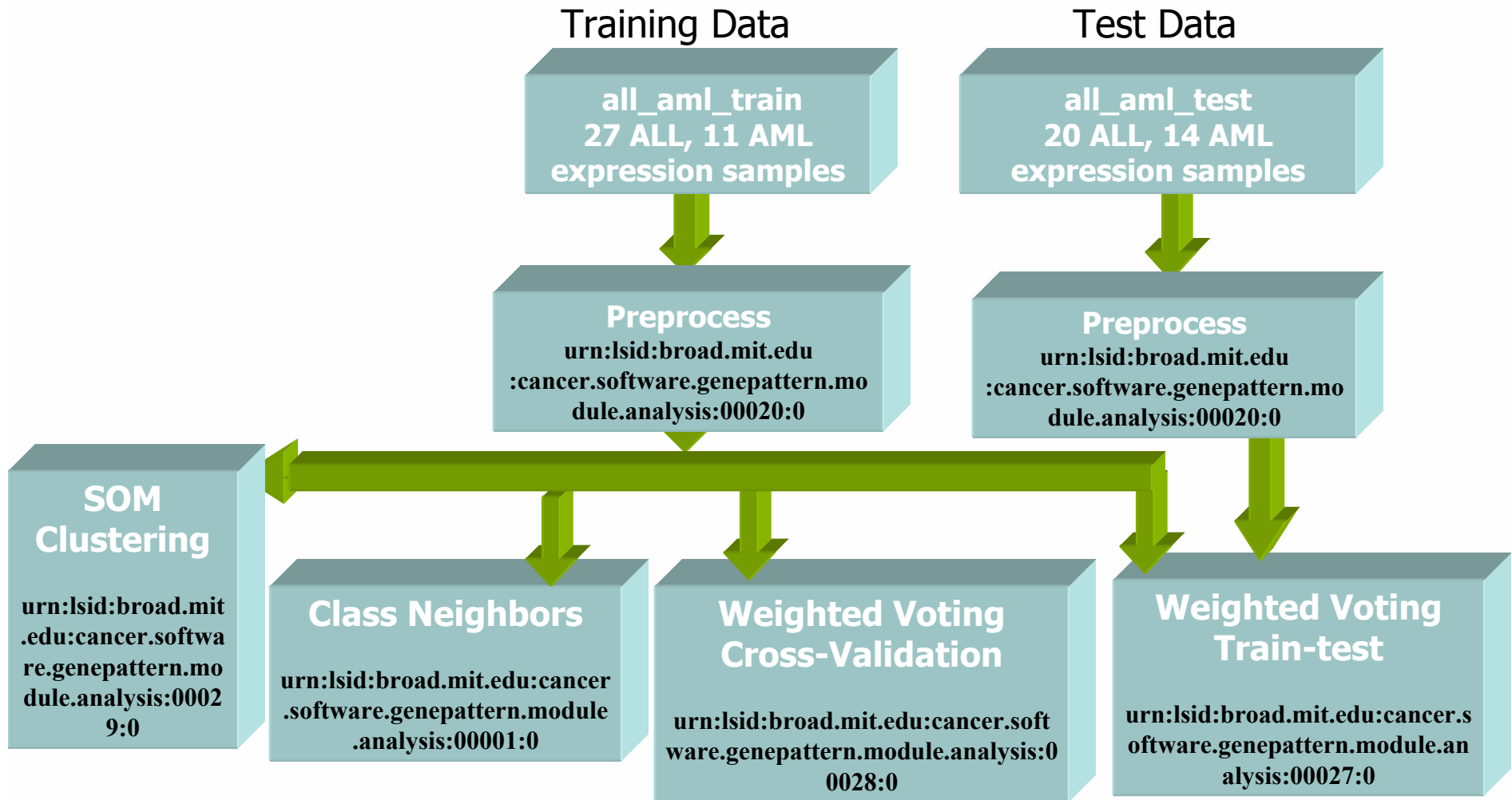
# How GenePattern is using LSIDs

1. Identify analysis tasks and pipelines via LSIDs
2. Create sharable pipelines referencing tasks via LSIDs
3. Provide a repository and retrieval for analysis tasks by LSID



Golub and Slonim et al., 1999

urn:lsid:broad.mit.edu:cancer.software.genepattern.module.pipeline:00001:0

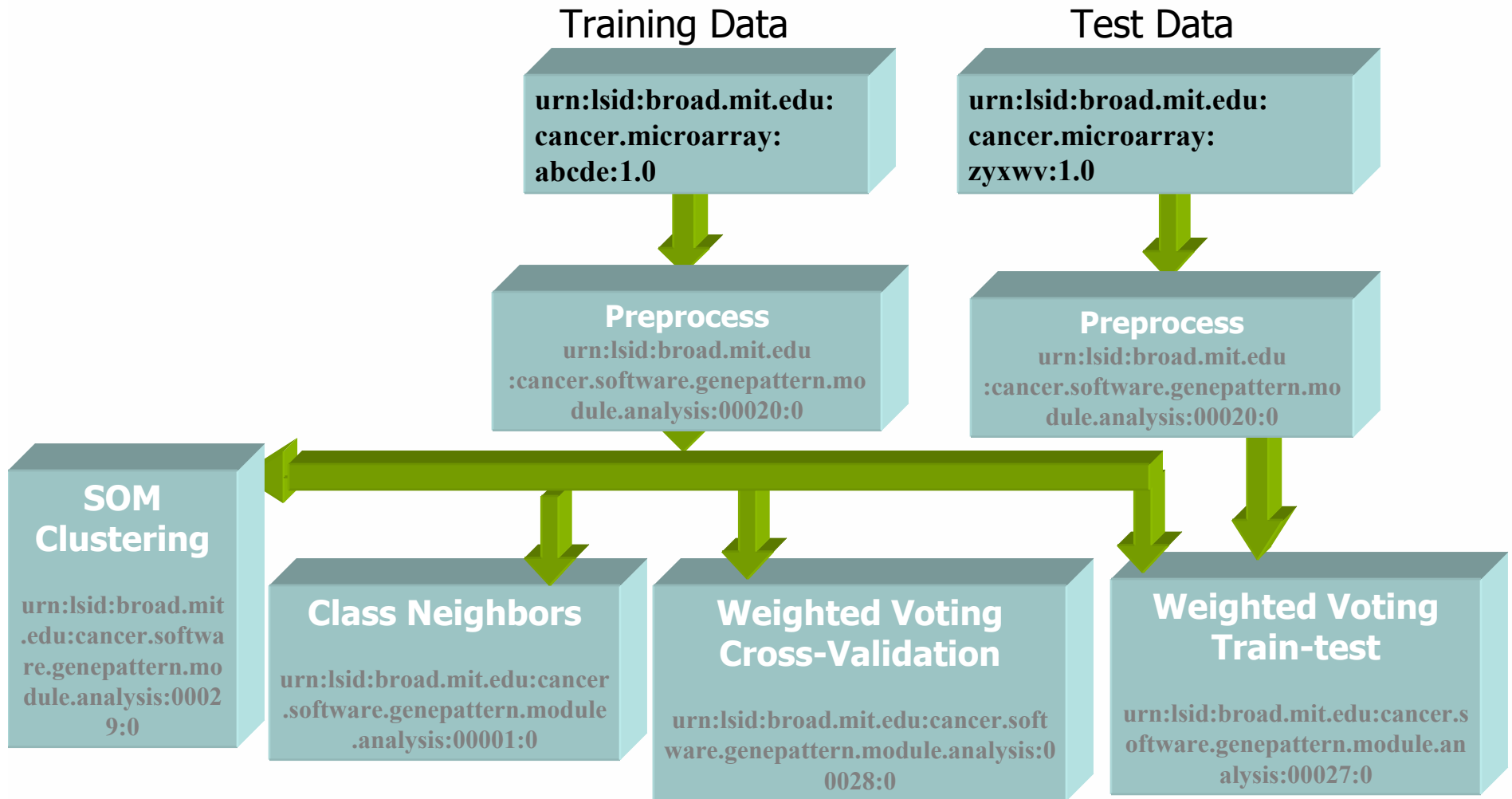


Golub and Slonim et al., 1999

- LSIDs enable
  - Reproducible research
    - exactly repeating an *in silico* experiment
  - ‘modernizing’ pipelines to latest
  - Tracking module provenance
- Someday
  - Data will be available via LSID too...



urn:lsid:broad.mit.edu:cancer.software.genepattern.module.pipeline:00001:0



Golub and Slonim et al., 1999

# Other LSID use at the Broad

## 1. Sample management

- Sharing samples (tissues, clones, etc) between program groups
- LSIDs identify samples
- Permits scientists to find all experiments done with a sample in any Broad program

# Other LSID use at the Broad

## 2. GeneCruiser web service

- annotation web service for microarray probes
- maps probe set identifiers to GO, GenBank, SwissProt etc
- Interface returns LSIDs to these other sources for their identifiers

# Use Cases and Future Directions

- What does it actually mean to identify a biological object such as "a gene"?
- How does LSID address structural elements of biological and chemical objects?
- What are the lessons learned from early implementations of LSID?

# Use Cases and Future Directions

- What granularity of object do we identify?
- Should LSID be a URI not a URN?
- Should virtual persistent identifiers for derived/calculated properties be used?
- What are the barriers to widespread use?
- Data/Metadata split – is this a problem?
  - Phil Lord mentioned @end of yesterday in MyGrid talk

## Best LSID quote...

- “LSIDs are in a sense just a sociological con trick, since they are nothing more than cheap and cheerful URNs” –David Shotten