Metadata Discussion @ COMBINE 2014

Samuel Friedman, Ph.D.

August 21, 2014

Center for Applied Molecular Medicine
Keck School of Medicine

University of Southern California

Background: 3-D simulation of a 1.7 mm tumor (cutaway view)

Keck School of
Medicine of USC

USC Center for Applied Molecular Medicine (CAMM)

MultiCellDS and Digital Cell Lines – MultiCellDS.org

MathCancer.org
@MathCancer
University of Southern California

Metadata Discussion Starters (A)

- All Metadata
  - Citation information (how to cite and credit data?)
    - Evidence Chain Ontology?
  - Database index: Where is the data nominally stored?
- Simulation Metadata – Likely Derived from Software
  - Software that generated this data
  - Who ran this code? When? What platform?
  - Timescales (File creation, Simulated Elapsed, Runtime)
- Experimental Metadata – Likely Derived from Microscopy
  - Experimental Conditions (Temperature, etc)
  - Who ran the experiment? When? Experimental Setup?
  - File creation time, Experiment Elapsed Time
  - Staining information (if any)
  - Image processing software information and who ran the pipeline
  - Links to previous set of files (e.g. original image file, intermediate files)

Metadata Discussion Starters (B)

- Digital Cell Lines Metadata
  - Cell Line. AKA What cells does this represent? MCF7, F9?
  - Where cells were ordered from?
  - Curation information – Community saying data is OK
    - Who curated? When was it last curated?
    - Previous versions? Links to newer versions?
- Digital Tissues (Down the road)
  - Links to original images
  - Who ran the pipeline?
  - Software used (e.g. image processing, algorithms)
  - Medical Annotations (e.g. case history, link to anonymous electronic health record)
- What elements should be required, optional, or recommended (i.e. worth a little effort)?

Keck School of Medicine of USC

USC Center for Applied Molecular Medicine (CAMM)

MultiCellDS and Digital Cell Lines – MultiCellDS.org

MathCancer.org
@MathCancer
University of Southern California