Internet Technology Group

The Semantic Layered Research Platform Prototype and KIM

Sean Martin

W3C Life Sciences Workshop

October 29, 2004

Researchers often have a highly manual research process and no integrated information management systems

Current Environment

- Vertical research methods and application silos
  - Researchers think in terms of discrete tools needed to complete each individual research step
  - Piece-meal single-task oriented apparatus
  - Researchers think of and treat IT like any other lab equipment
- Highly skilled, highly manual processes
- Collaboration and community communication occur through word of mouth, conferences, published papers, and email
- Information often mislaid, lost, or dots not connected, context not trapped
- Or inordinate amount of time is spent manually keep track of information
- Increasing recognition that In Silico, or digital biology, is the way forward

Researchers often have a highly manual research process and no integrated information management systems

Entire endeavor is about creating information, but while it is created, it is not managed

- Time wasted manually moving data from one place to another
- Time spent checking and integrating data from different silos
- Information is lost
- In silico is treated like the physical world without seeing the power of data connections
- Collaboration is far more difficult
- Scientists are isolated from key data

This lack of an information integration backbone results in an inefficient research process

Semantic Layer Research Platform is designed to improve research process by providing an integrated information management system

Reduce barriers to HPC/GRID computing

- "Cost" (time/manual setup) of running/rerunning experiments is slashed
  - Compute requirements of researchers increase dramatically as manual setup declines
- SLRP tracks, integrates and aids in all aspects of a research project, including data generated from experiment and literature searches with computational experiments, collaborations, and publication
- Highly repeatable experiments, better provenance
  - Clear Information trail
  - Exact algorithms, Exact data
  - Able to reference/send entire context to colleagues
- Better collaboration
  - Collaborative annotation
  - Seamless User Interfaces
  - Easier peer review of experiments (and their own "riffs")

The Semantic Layer Research Platform requires new technologies and new uses of existing technologies

- Cart — Semantic data infrastructure
- Slingshot — BPAL-like workflow capabilities for the Semantic Web
- DDR — Write once, read-only object repository
- Jastor — Code generation for RDF handling
- Telar — Using RDF in Eclipse
- LSID — System underpinning: All objects and their relationships uniquely named

- Standards-based — RDF, LSID, CIM, WSDL, SOAP, etc.
- Open source

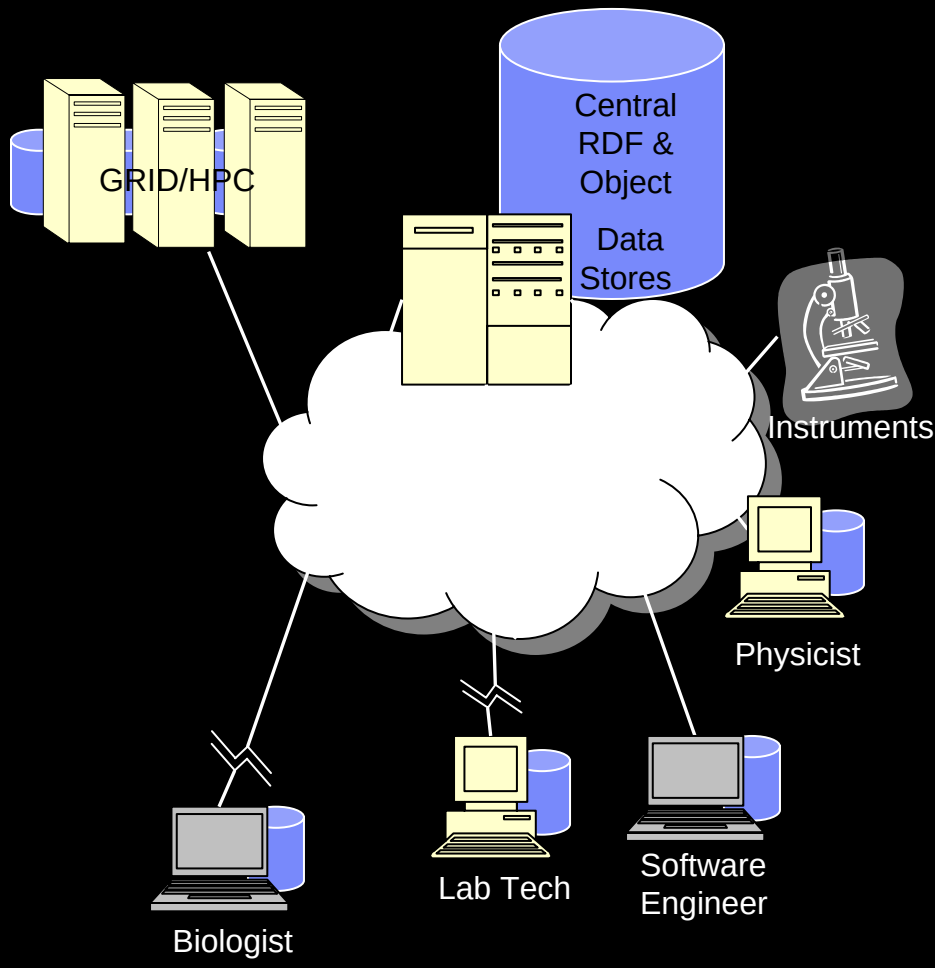
Adding up to the Semantic Layer Research Platform

CART: Semantic Web Infrastructure for the Enterprise

Collections, Access Control, Revision Tracking

CART is a Client / Server RDF storage architecture built on the WebSphere Platform (DB2, EventBroker)

- Current situation:
  - Simple RDF models published on web
  - Client applications store metadata in RDF
- For the Enterprise we need:
  - Multi-user
    - Distributed replicated/synchronized views
    - Distributed Events
    - Transaction support
    - Access Control
  - Reification
  - Duplicate triple support
  - Collections
  - Scalability
  - History / Auditable
  - Data federation



Slingshot

Enables dynamic semantic workflows

- Problem: Scientists run dynamic, “ad-hoc” workflows of tasks that execute in heterogeneous environments
  - Workflows are developed quickly and modified often
  - Workflows are carried out cooperatively in several types of participants including HPC/GRID applications, Web Service invocations and user-interactive client applications
- Solution: OWL-S++ execution engine
  - ++ == Groundings for GRID/HPC and workspace execution
  - Current state of flow stored in CART (central RDF-store)
  - Participants are notified when a new flow is submitted to the store and when the state of the flow changes (PUB-SUB)
  - Each participant independently evaluates the flow to determine if it has to contribute anything at the moment
  - When a participant commits a task, the state of the flow changes and all participants must re-evaluate

Distributed Data Repository

Write once, read only file object store that provides secure access by LSID reference to data over AHPFS in grid clusters, WAN links, SOAP, and HTTP

- All aspects of DDR are described in RDF using OWL ontologies
- Repository based on Linux and ReiserFS with extended file system attributes
- Server stack based on Apache with mod_perl providing access via SOAP and HTTP
- Remote administration of repository via XML/RDF in SOAP
- Granular access control via file system ACLs
- Java client API, including LSID protocol handler, provides online/offline writing, reading, and synchronizing of data and associated metadata

Jastor

Code generation tool for creating custom libraries

- Developers don't actually want to deal with RDF
- Provides object-oriented view, including events, of RDF
  - Conforms to a particular OWL ontology
- Extensible to any object oriented programming language
  - Java classes
- OWL
  - Compliant with OWL-Lite
  - Supports an increasing number features of OWL-DL and OWL-Full

Life Science Identifiers (LSID)

URN naming, discovery and access standard for data and metadata

- All system & data objects and concepts have a LSID URI
  - CART identifies documents/collections with LSIDs
  - LSIDs are used to name ontologys, formats & predicates
  - DDR provides access to files via LSID
- CVS <-> LSID Resolution service to track code modules
- Use LSIDs to track and access external references wherever possible

Telar is a framework for building Semantic user interfaces

A framework designed to help developers build user interfaces capable of weaving data from a myriad of data sources into a single application for creating, exploring and organizing information

Java library for Eclipse

Enables mapping of dynamic RDF data to SWT widgets

Groups widgets into logical information clusters and manages their life cycle for UI forms presentation

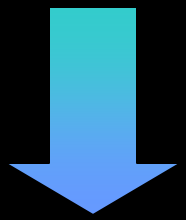
Maps onto CART client-side API for synchronization support

Problems:

- Information about any subject cannot be contained within a single database or structure
- Users need a larger view of the information available on any given subject
- The number of applications users need to deal with all of today's data structures is overwhelming
- Existing applications are generally hard-coded to perform a limited set of functions on a few data structures

Interdisciplinary Work

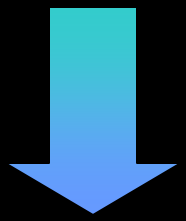
- **Scientific Diversity:** **Multiple disciplines, methods and techniques**
- **Geographic Distance:** **Multiple sites (academia, government, corporate)**
- **Data Infrastructure:** **Multiple distributed data bases (data sharing and integration, data safety and confidentiality)**



Knowledge-Integrated Modeling (KIM)

Knowledge-Integrated Modeling (KIM)

- Integrates peer-reviewed biomedical data into a 3D computational modeling, simulation and visualization platform (within a GRID environment).
- Enables large-scale biomedical collaborations.
- Guides experimental data acquisition and specifies data-base mining.
- Ultimately, is intended for both research *and* clinical use, i.e. for diagnostic purposes and treatment outcome predictions.

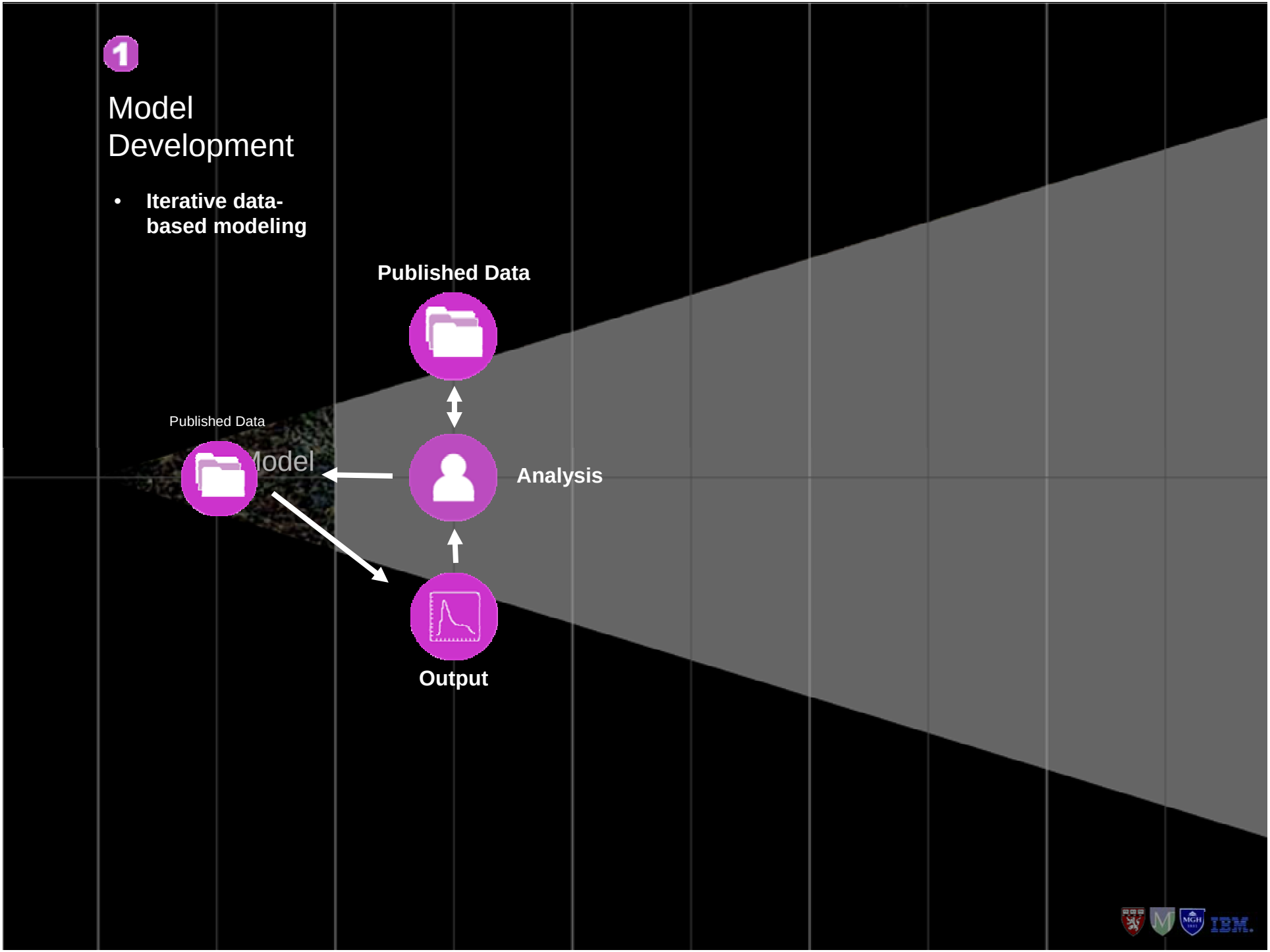


5 Milestones

1

Model Development

- **Iterative data-based modeling**



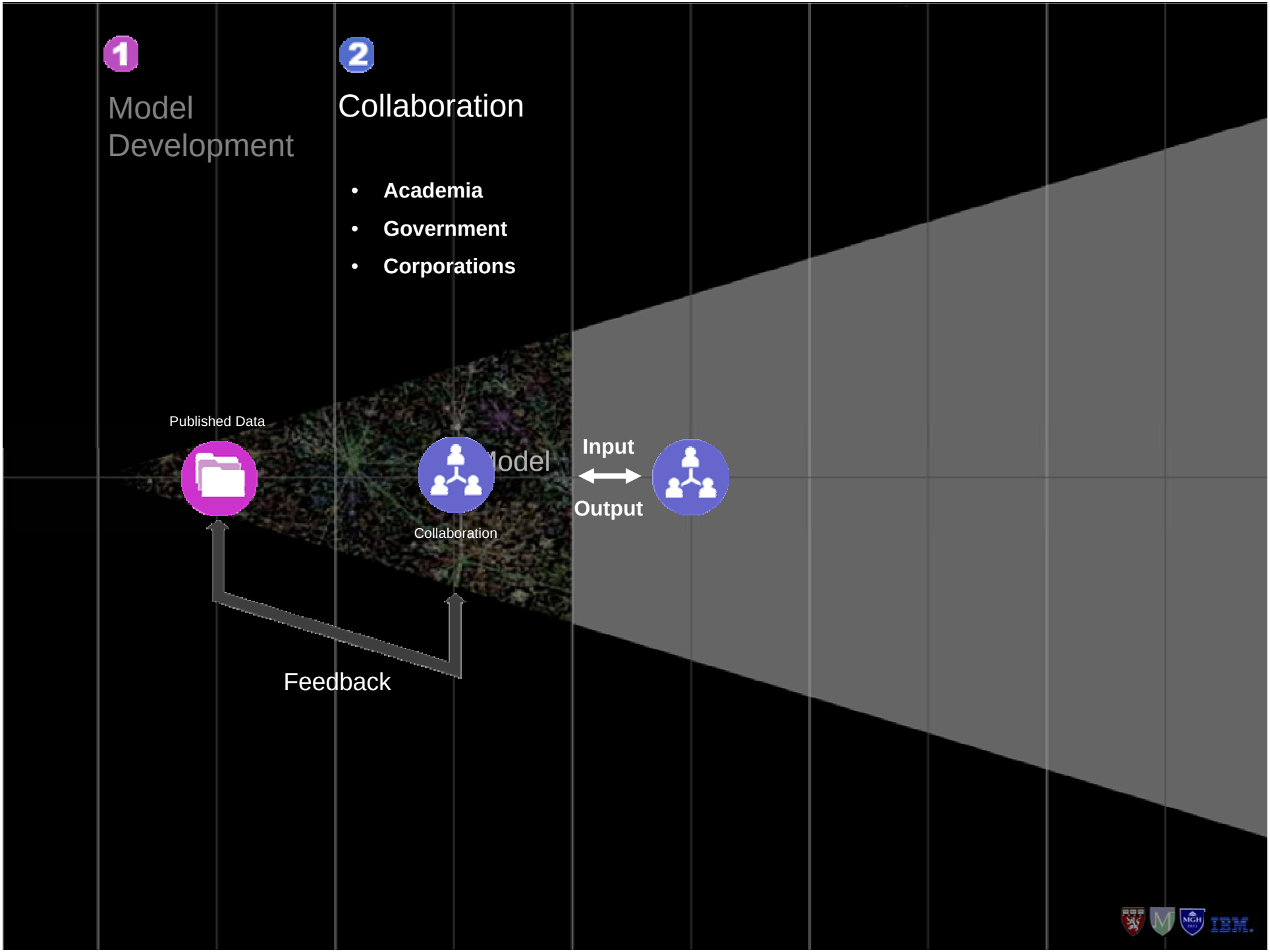
1

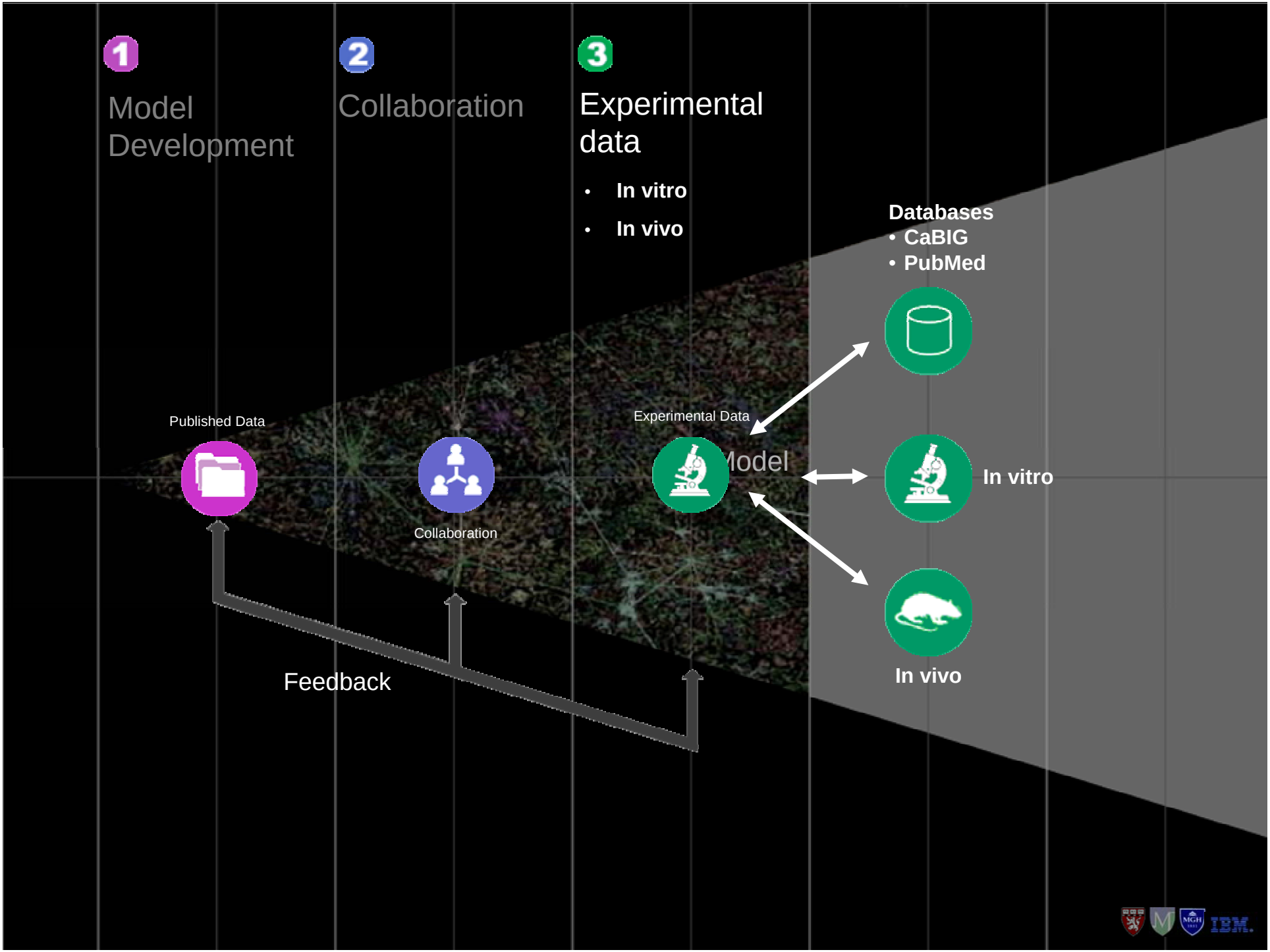
Model Development

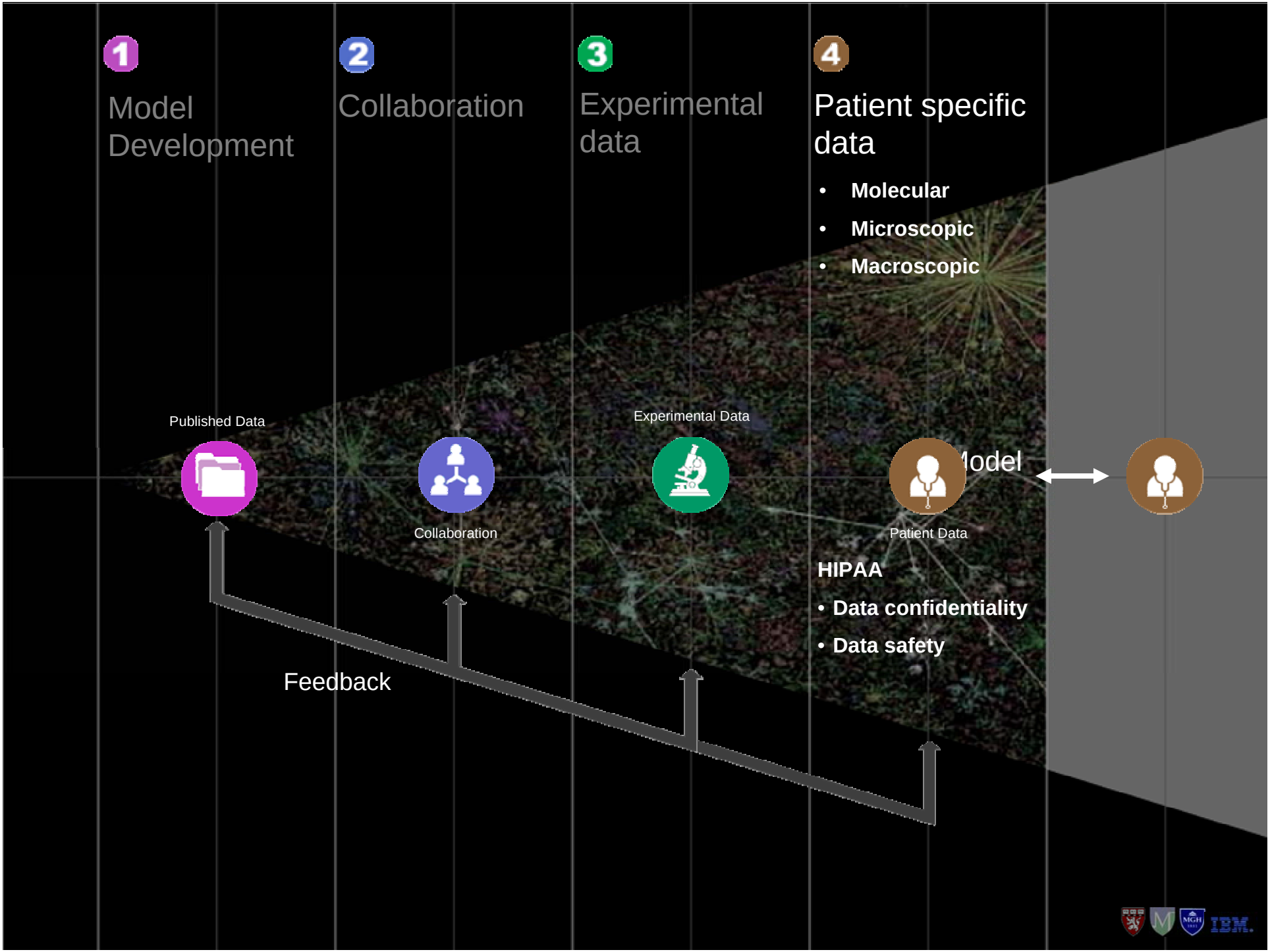
2

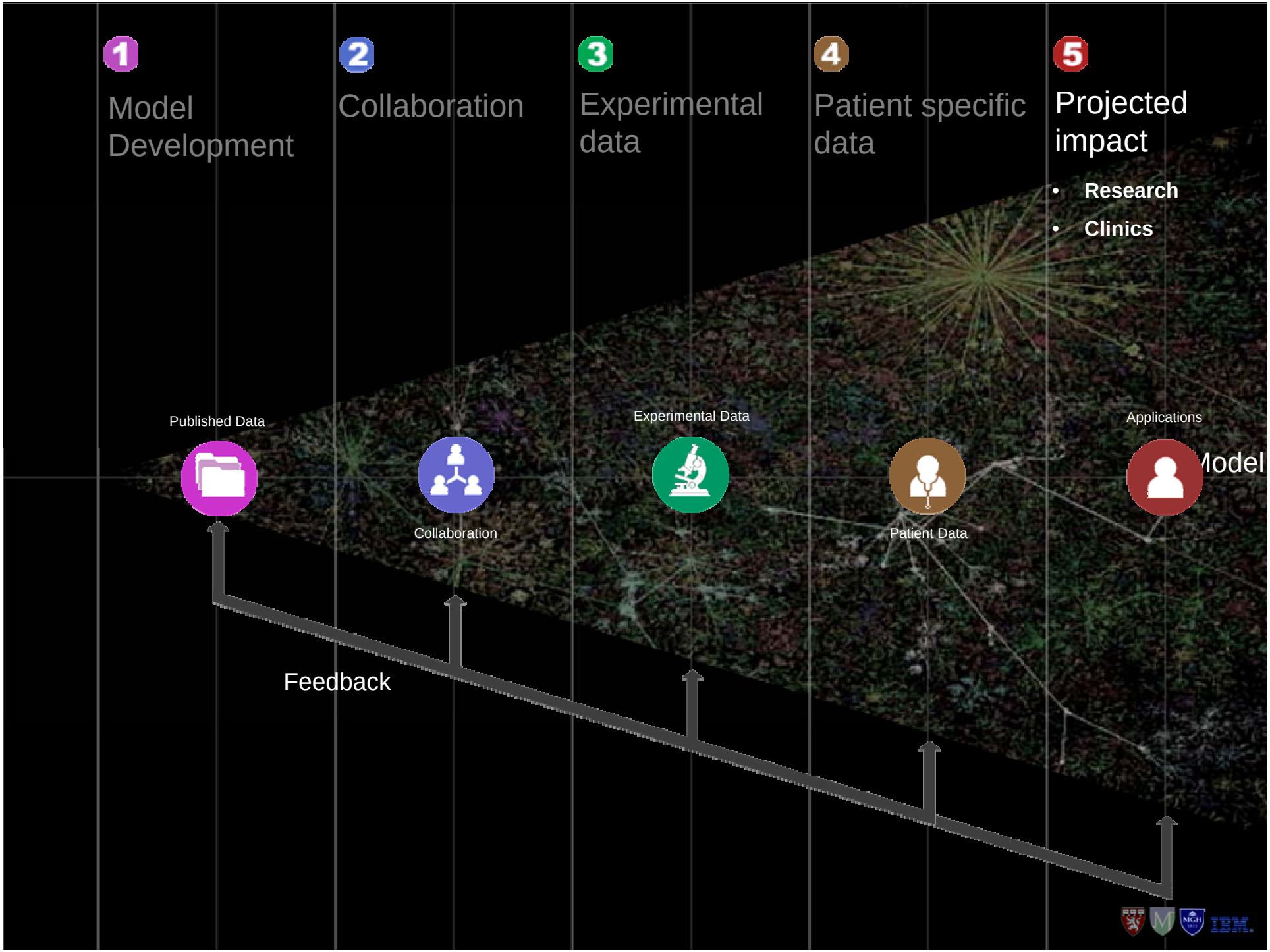
Collaboration

- **Academia**
- **Government**
- **Corporations**



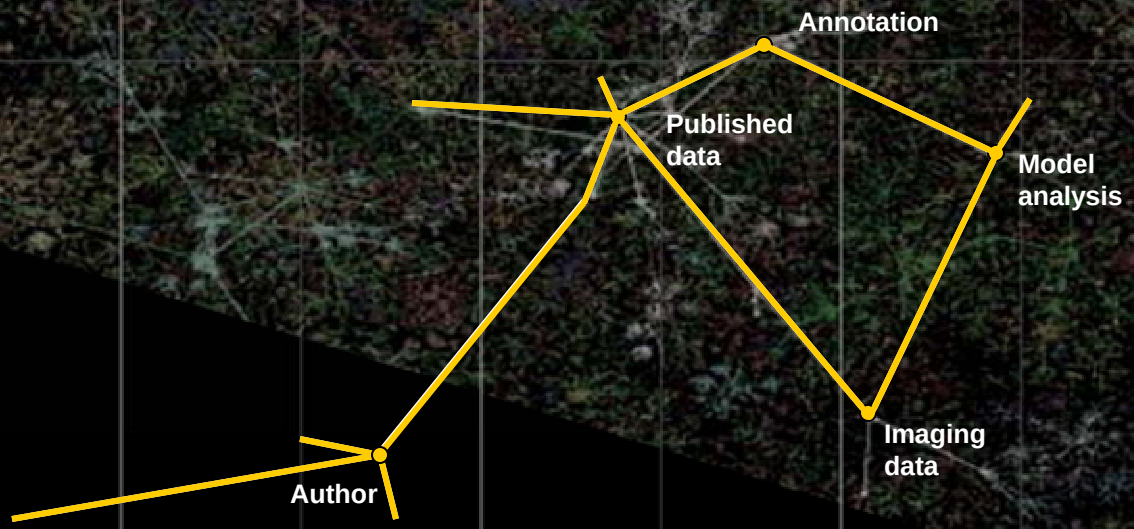

1
Model Development
2
Collaboration
3
Experimental data
• In vitro
• In vivo
Databases
• CaBIG
• PubMed
Published Data
Collaboration
Experimental Data
Model
In vitro
In vivo
Feedback


1
Model Development
2
Collaboration
3
Experimental data
4
Patient specific data
• Molecular
• Microscopic
• Macroscopic
Published Data
Collaboration
Experimental Data
Patient Data
HIPAA
• Data confidentiality
• Data safety
Feedback


1
Model Development
2
Collaboration
3
Experimental data
4
Patient specific data
5
Projected impact
• Research
• Clinics
Published Data
Collaboration
Experimental Data
Patient Data
Applications
Model
Feedback

Semantic Layer

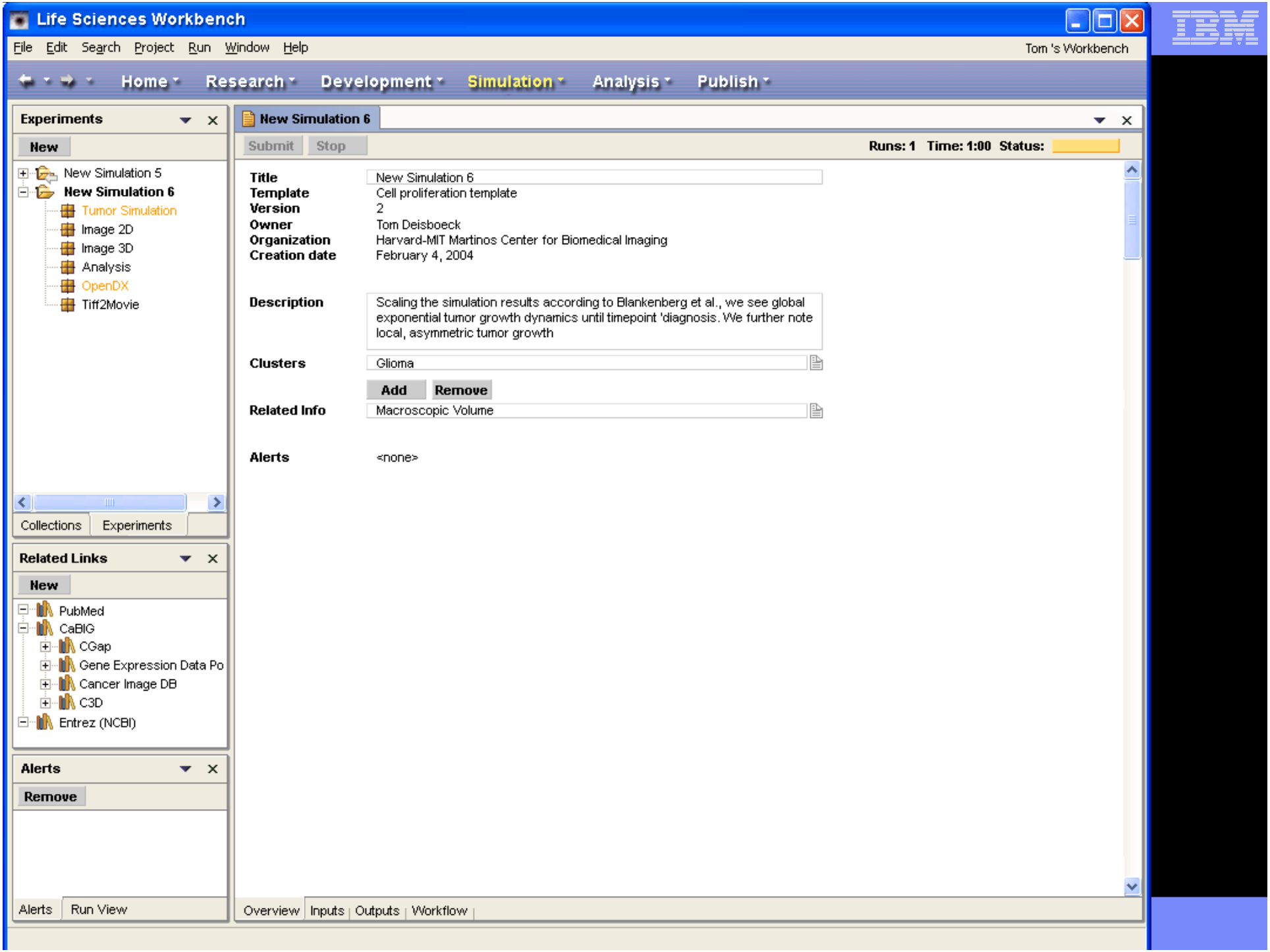
- Rich network of objects and relationships
- Capture of information contexts in metadata
- Enabler of systems biology

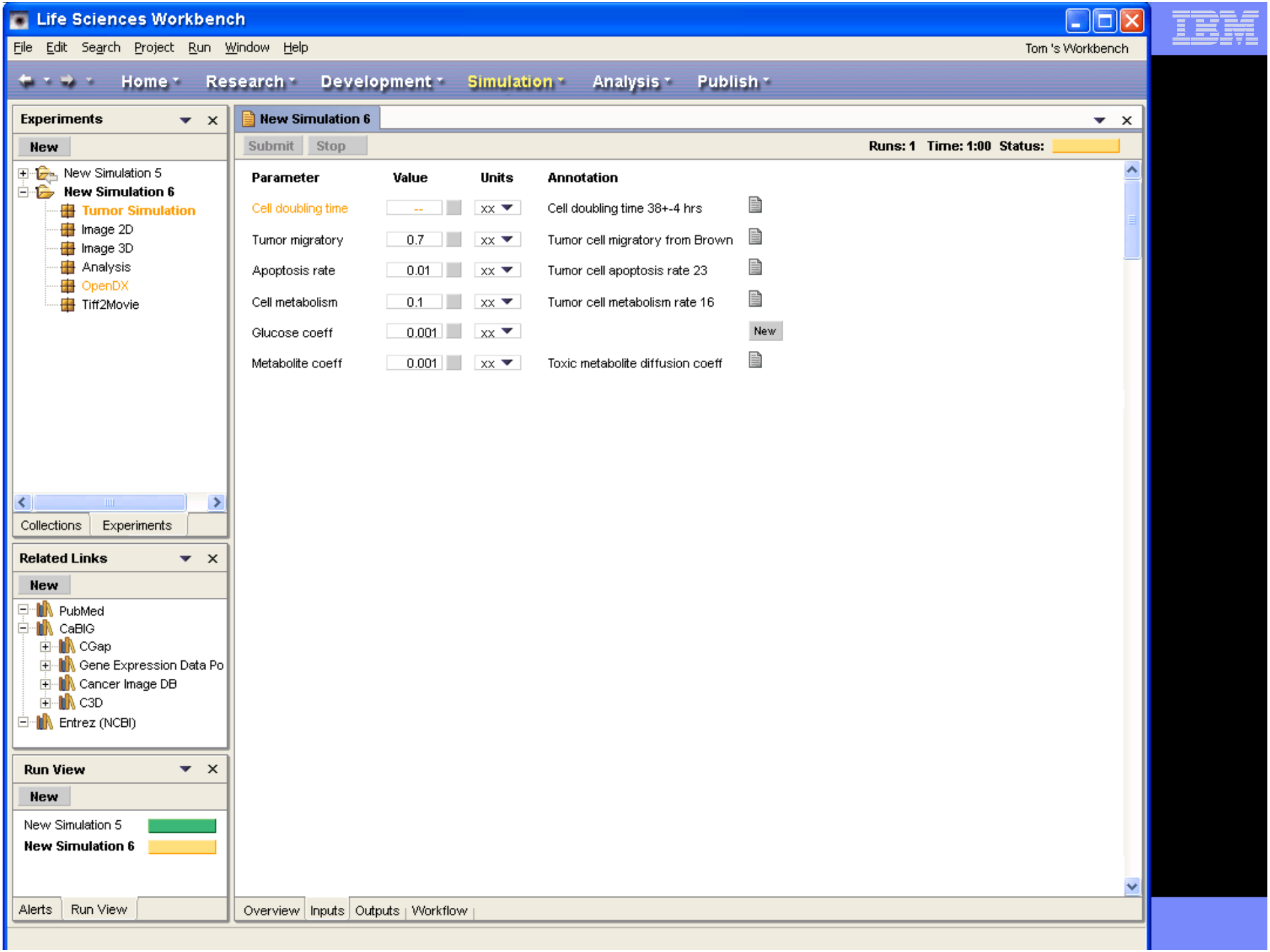


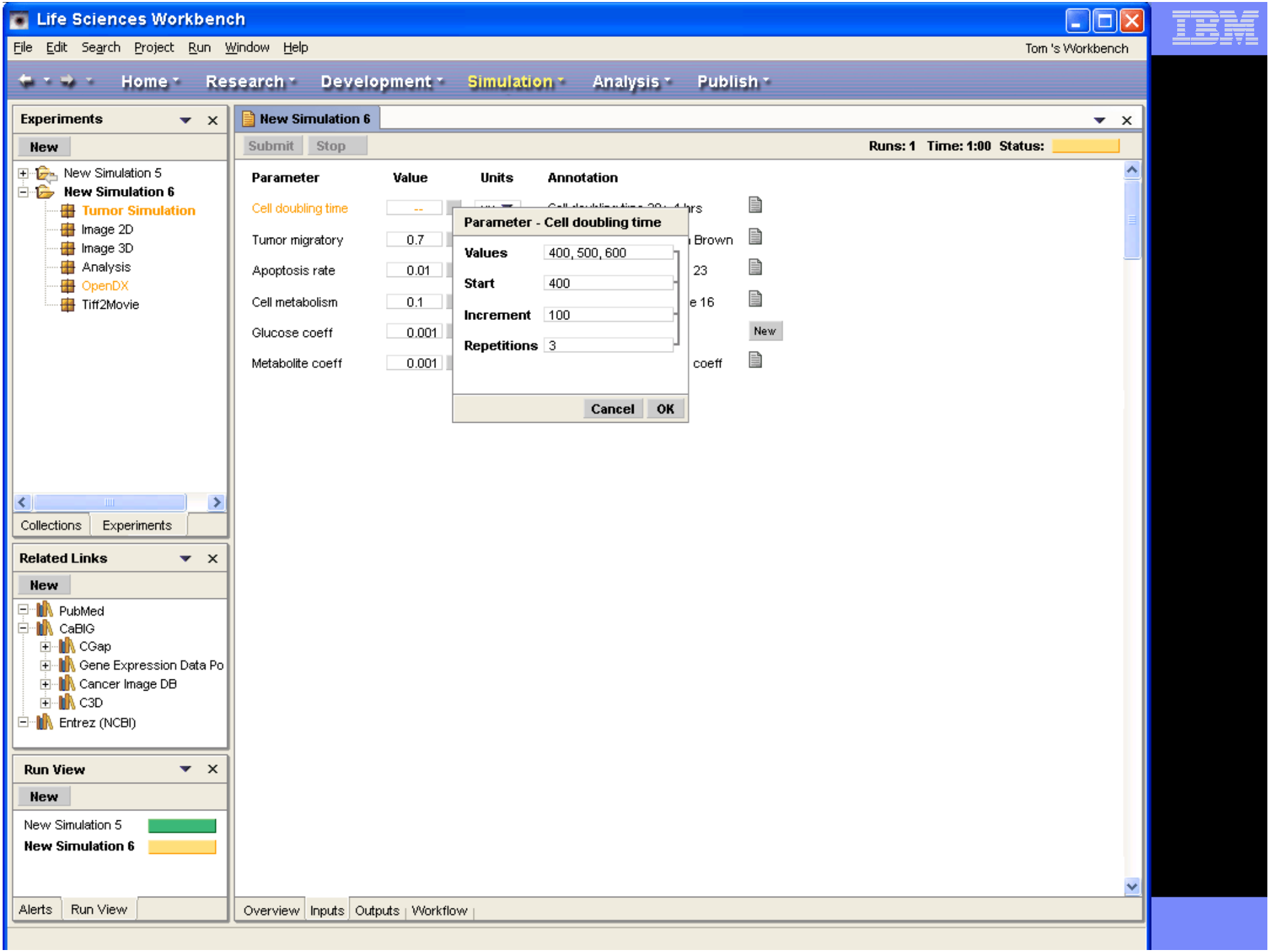
Usability and Scenarios

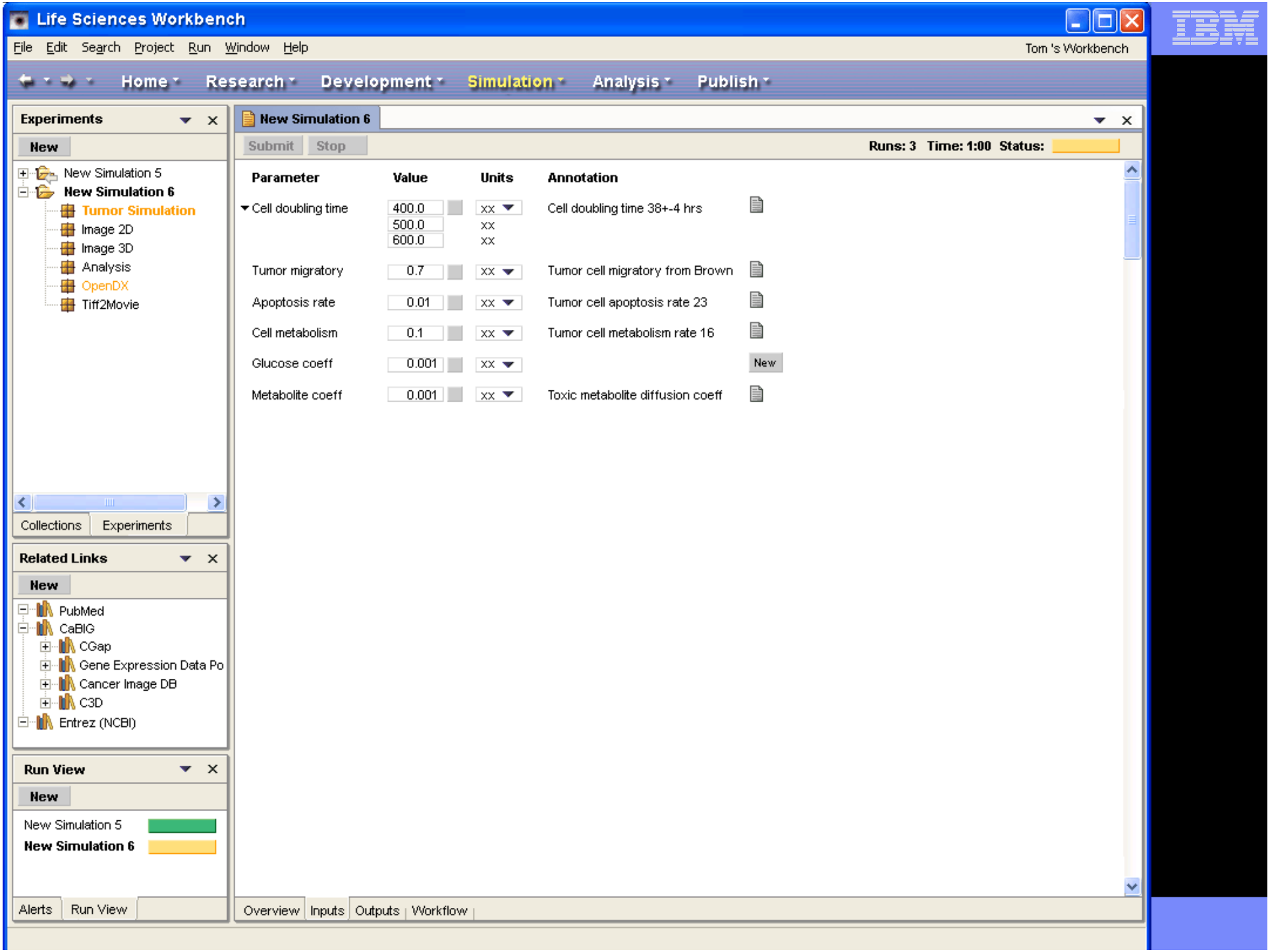
- Observed and interacted with customers
  - Identified workbench personas
  - Understood user activities and tasks
- Created scenarios based on observed working practices of scientists in their various personas
  - Biologists, physicists, informaticians
- Created user interface designs

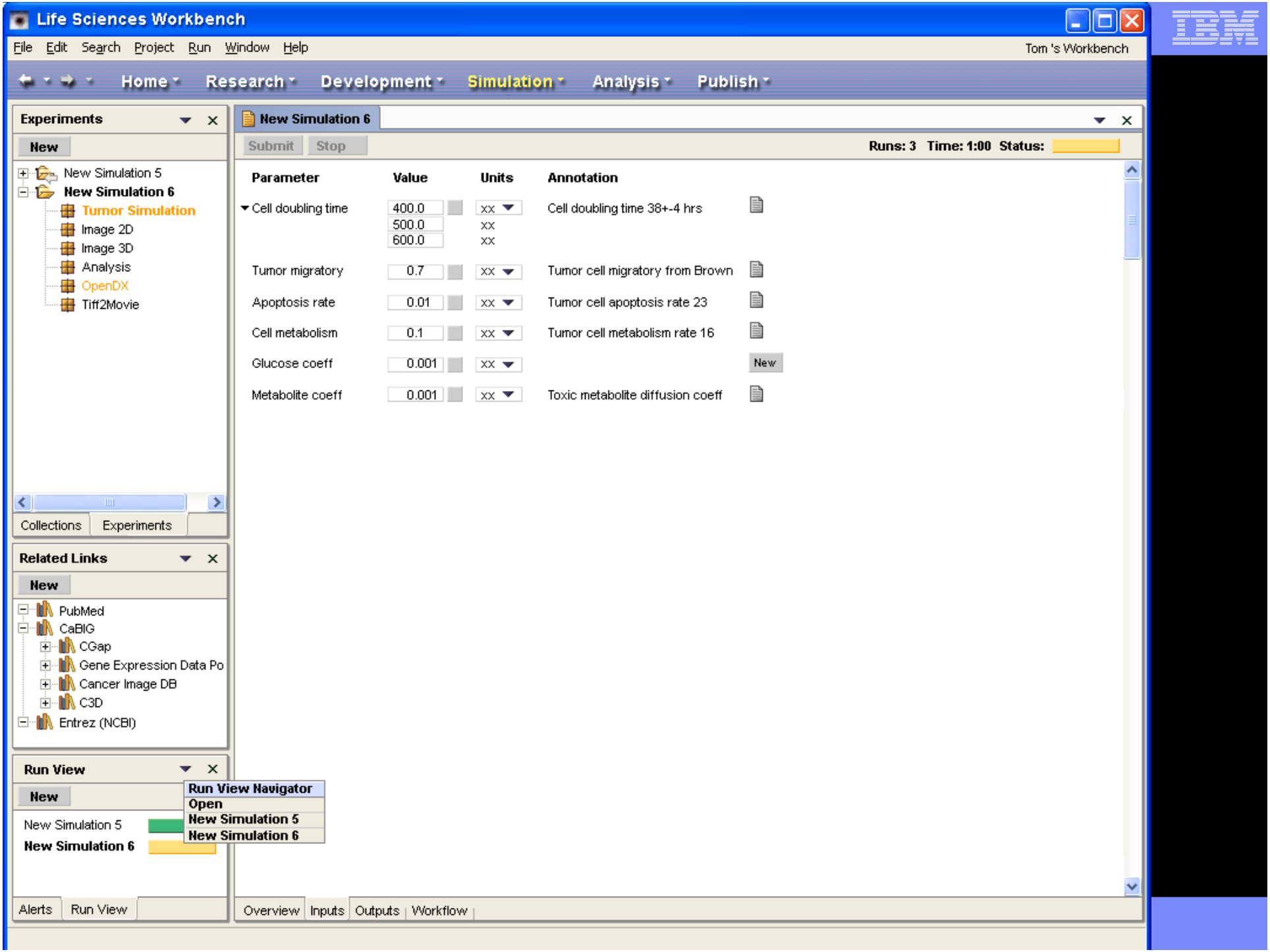
Scenario: Creating simulation

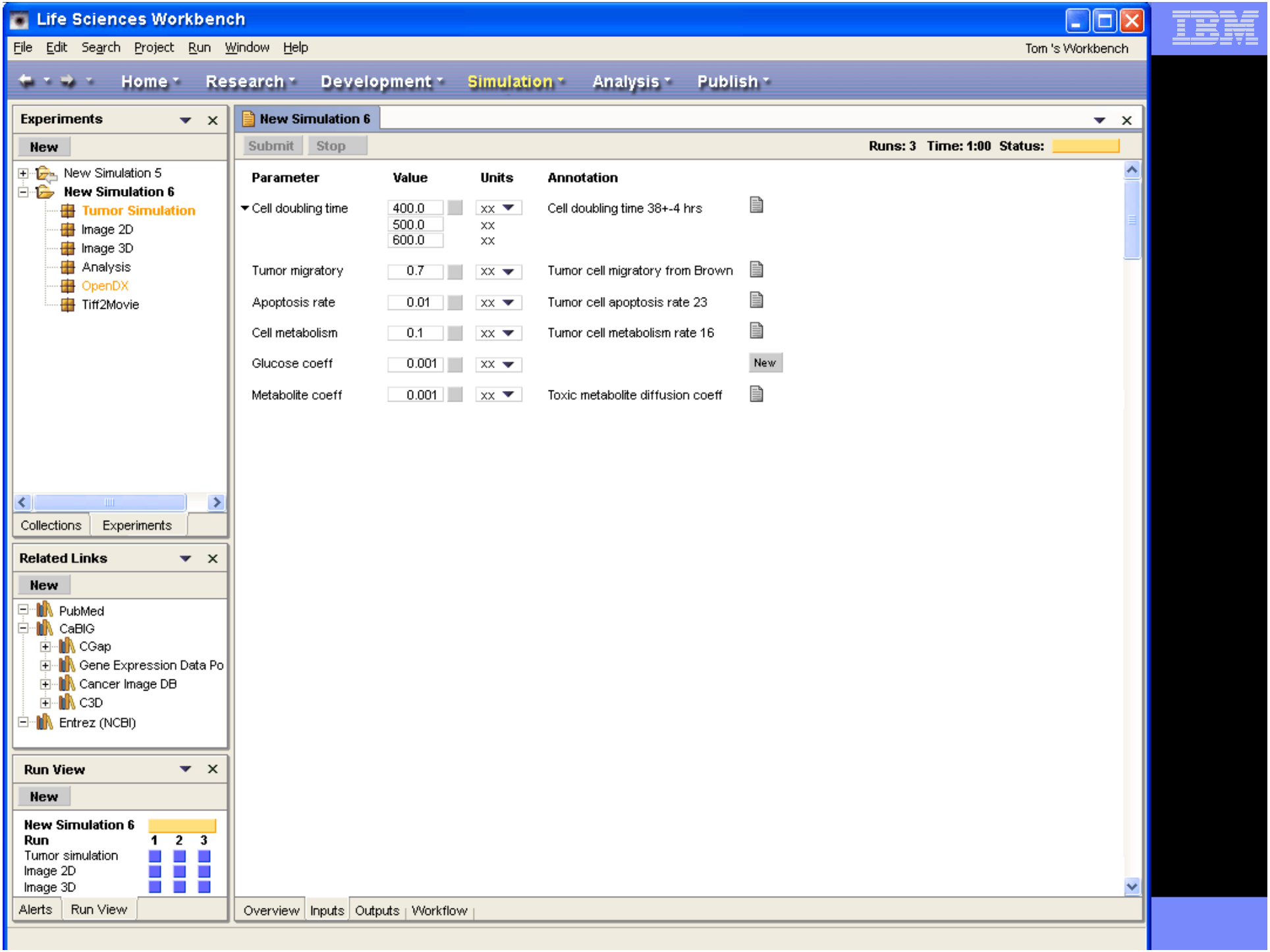

Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Alerts
Remove
Alerts Run View
New Simulation 6
Submit Stop
Runs: 1 Time: 1:00 Status:
Title New Simulation 6
Template Cell proliferation template
Version 2
Owner Tom Deisboeck
Organization Harvard-MIT Martinos Center for Biomedical Imaging
Creation date February 4, 2004
Description Scaling the simulation results according to Blankenberg et al., we see global exponential tumor growth dynamics until timepoint 'diagnosis. We further note local, asymmetric tumor growth
Clusters Glioma
Add Remove
Related Info Macroscopic Volume
Alerts <none>
Overview Inputs Outputs Workflow

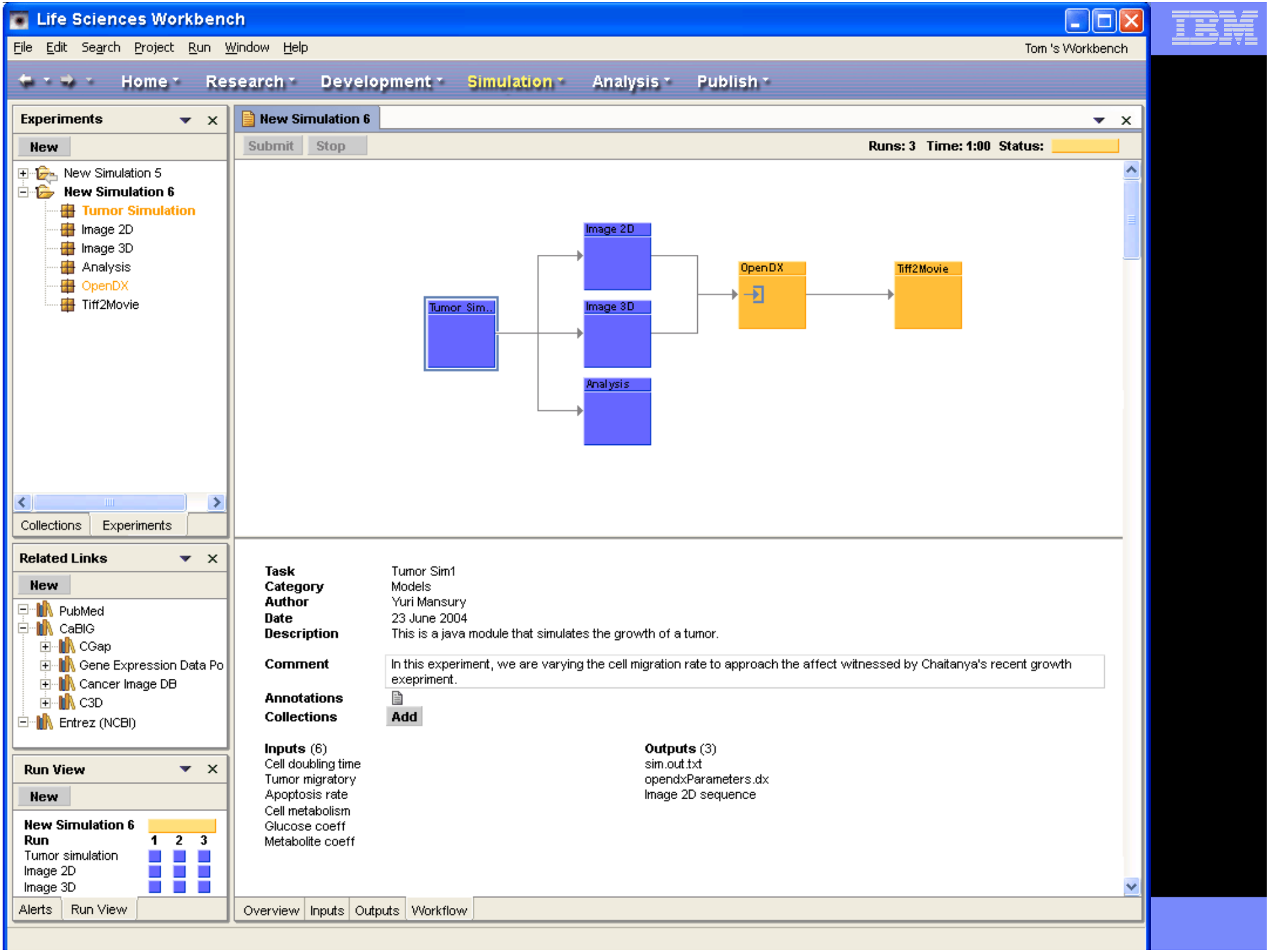

Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 5
New Simulation 6
Alerts Run View
New Simulation 6
Submit Stop
Runs: 1 Time: 1:00 Status:
Parameter Value Units Annotation
Cell doubling time -- xx Cell doubling time 38+-4 hrs
Tumor migratory 0.7 xx Tumor cell migratory from Brown
Apoptosis rate 0.01 xx Tumor cell apoptosis rate 23
Cell metabolism 0.1 xx Tumor cell metabolism rate 16
Glucose coeff 0.001 xx New
Metabolite coeff 0.001 xx Toxic metabolite diffusion coeff
Overview Inputs Outputs Workflow

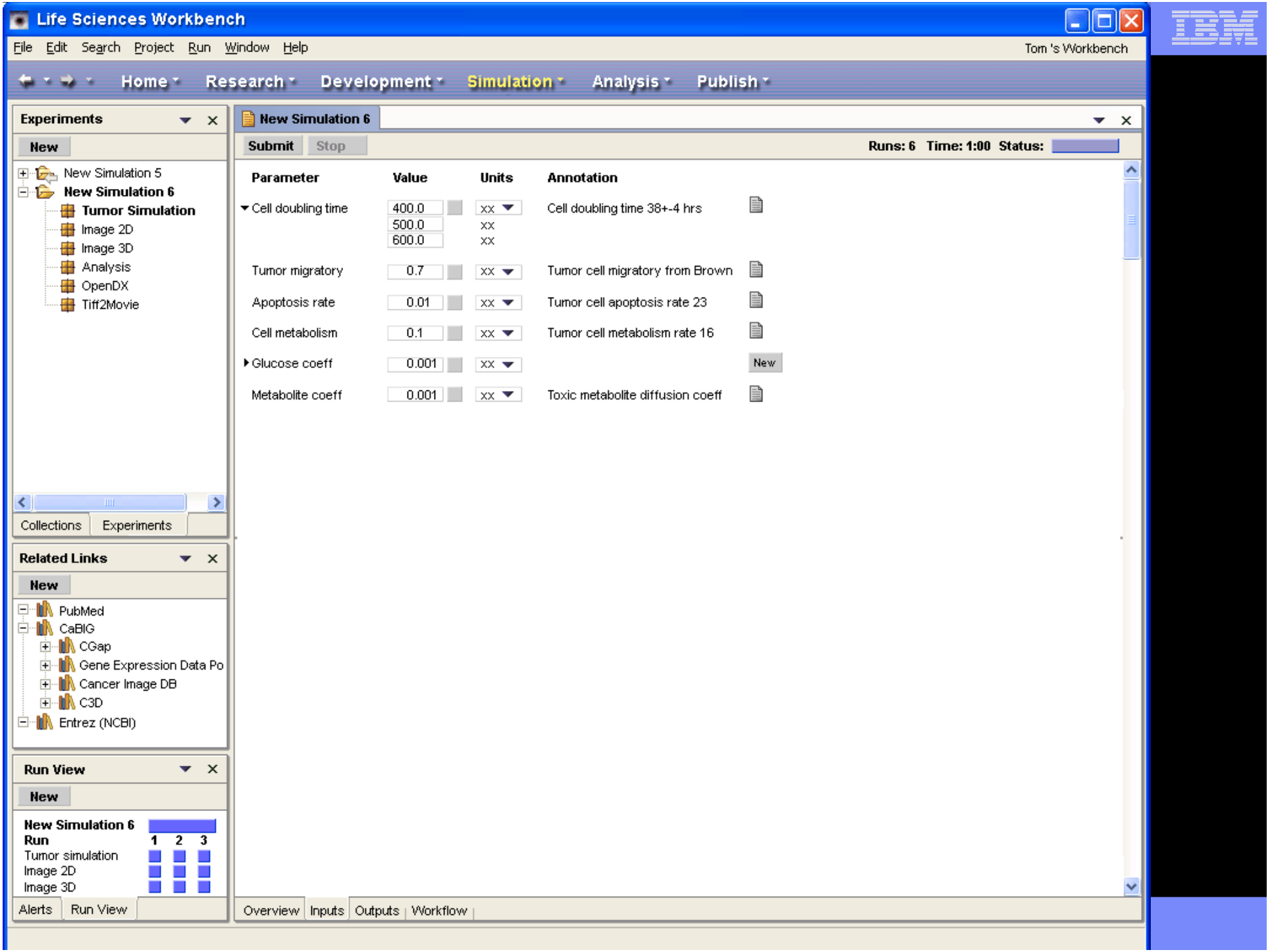

Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 5
New Simulation 6
Alerts Run View
New Simulation 6
Submit Stop
Runs: 1 Time: 1:00 Status:
Parameter Value Units Annotation
Cell doubling time --
Tumor migratory 0.7
Apoptosis rate 0.01
Cell metabolism 0.1
Glucose coeff 0.001
Metabolite coeff 0.001
Parameter - Cell doubling time
Values 400, 500, 600
Start 400
Increment 100
Repetitions 3
Cancel OK
New
Overview Inputs Outputs Workflow

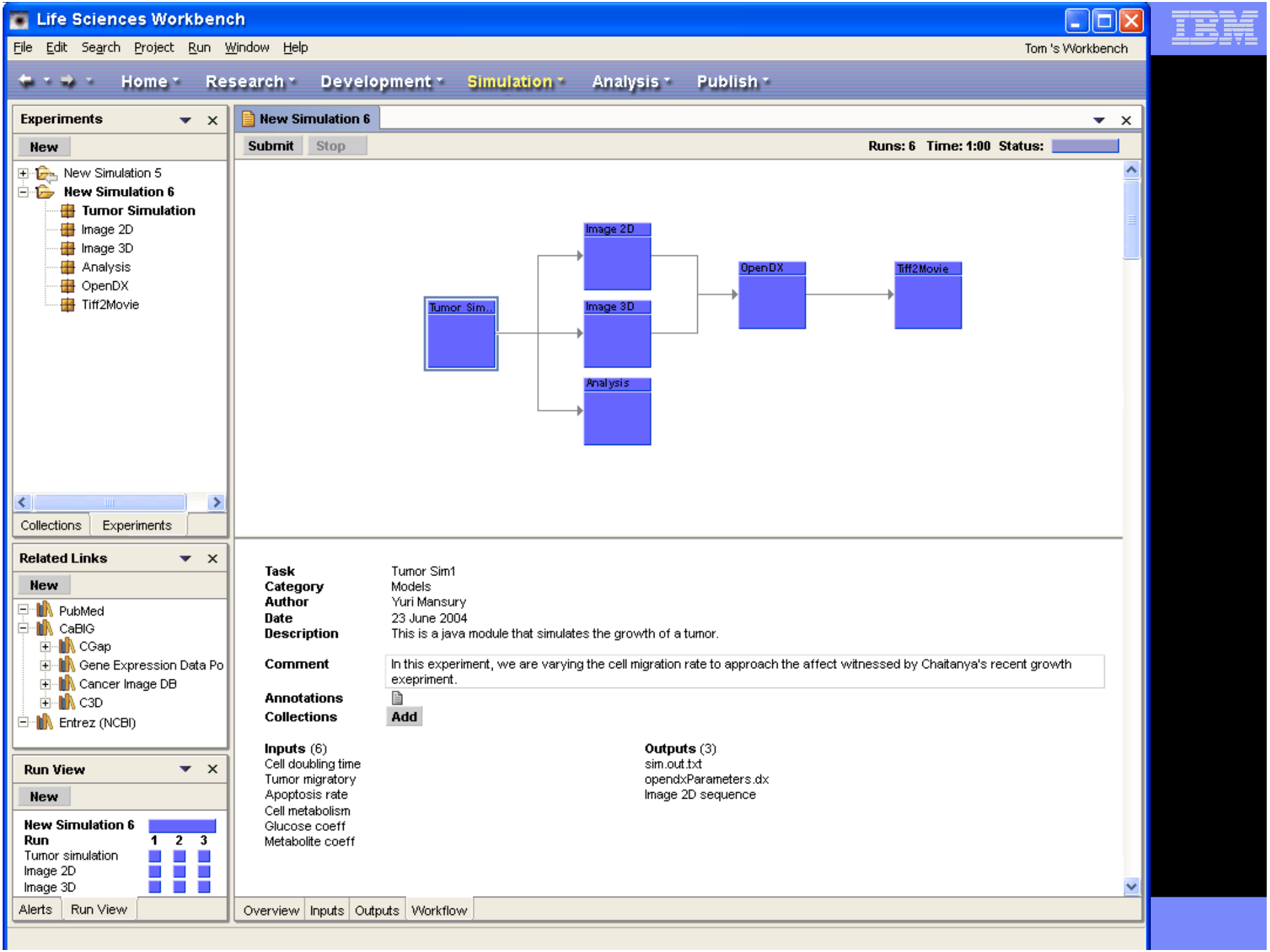

Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 5
New Simulation 6
Alerts Run View
New Simulation 6
Submit Stop
Runs: 3 Time: 1:00 Status:
Parameter Value Units Annotation
Cell doubling time 400.0 xx Cell doubling time 38+-4 hrs
500.0 xx
600.0 xx
Tumor migratory 0.7 xx Tumor cell migratory from Brown
Apoptosis rate 0.01 xx Tumor cell apoptosis rate 23
Cell metabolism 0.1 xx Tumor cell metabolism rate 16
Glucose coeff 0.001 xx New
Metabolite coeff 0.001 xx Toxic metabolite diffusion coeff
Overview Inputs Outputs Workflow

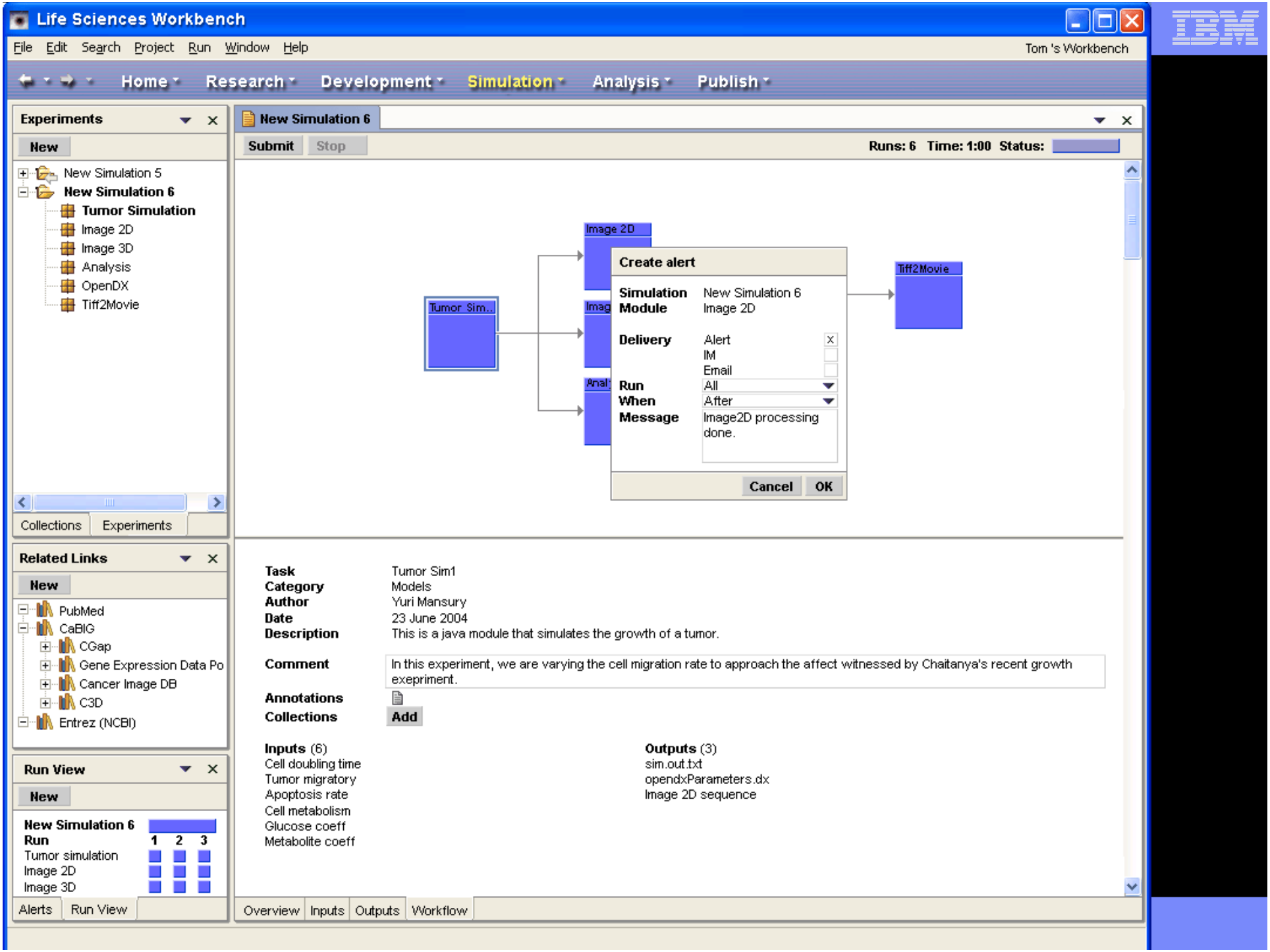

Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 5
New Simulation 6
Run View Navigator
Open
New Simulation 5
New Simulation 6
Alerts Run View
New Simulation 6
Submit Stop
Runs: 3 Time: 1:00 Status:
Parameter Value Units Annotation
Cell doubling time 400.0 xx Cell doubling time 38+-4 hrs
500.0 xx
600.0 xx
Tumor migratory 0.7 xx Tumor cell migratory from Brown
Apoptosis rate 0.01 xx Tumor cell apoptosis rate 23
Cell metabolism 0.1 xx Tumor cell metabolism rate 16
Glucose coeff 0.001 xx New
Metabolite coeff 0.001 xx Toxic metabolite diffusion coeff
Overview Inputs Outputs Workflow

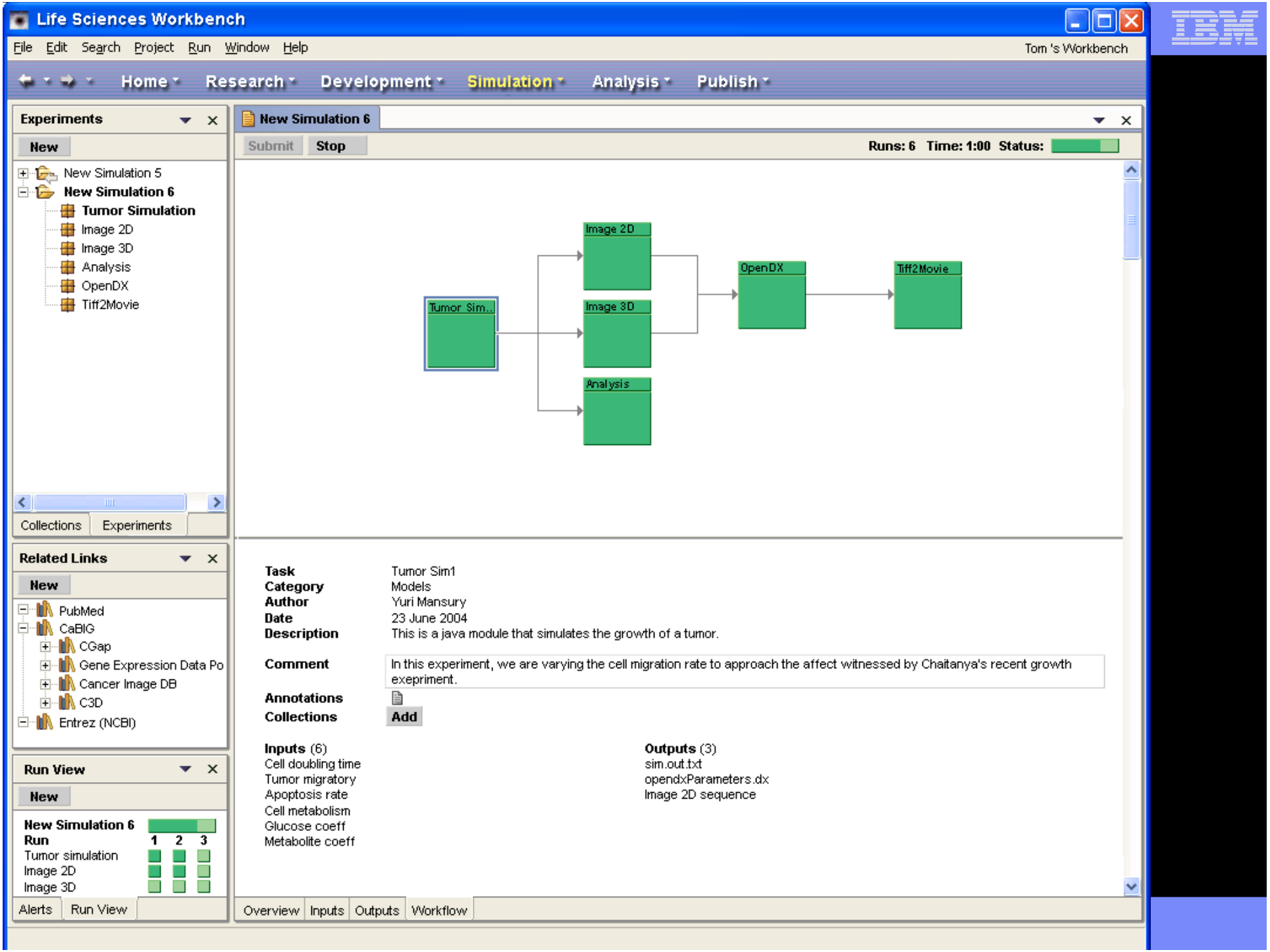

Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 6
Run 1 2 3
Tumor simulation
Image 2D
Image 3D
Alerts Run View
New Simulation 6
Submit Stop
Runs: 3 Time: 1:00 Status:
Parameter Value Units Annotation
Cell doubling time 400.0 xx Cell doubling time 38+-4 hrs
500.0 xx
600.0 xx
Tumor migratory 0.7 xx Tumor cell migratory from Brown
Apoptosis rate 0.01 xx Tumor cell apoptosis rate 23
Cell metabolism 0.1 xx Tumor cell metabolism rate 16
Glucose coeff 0.001 xx New
Metabolite coeff 0.001 xx Toxic metabolite diffusion coeff
Overview Inputs Outputs Workflow

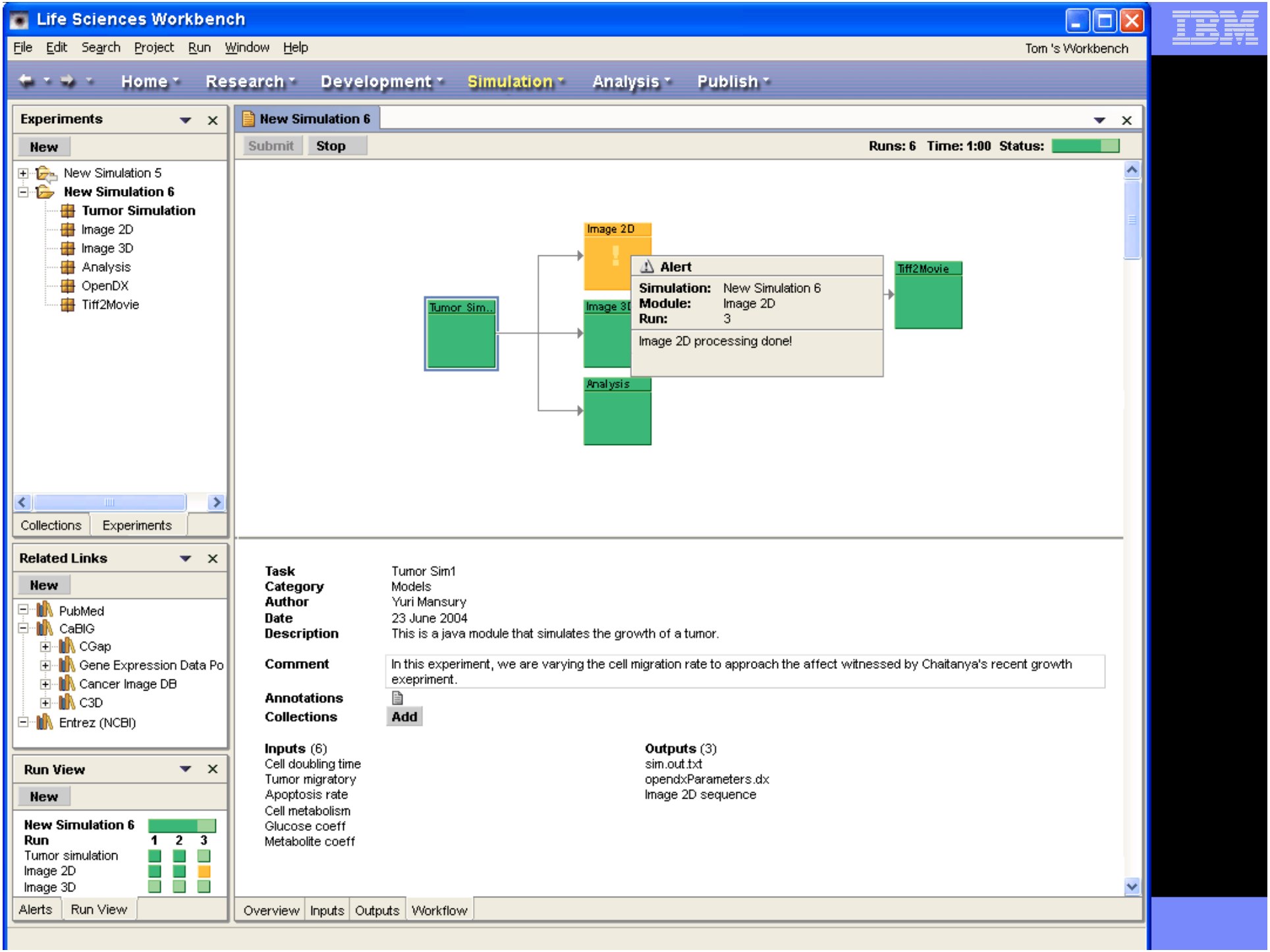

Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 6
Run 1 2 3
Tumor simulation
Image 2D
Image 3D
Alerts Run View
New Simulation 6
Submit Stop
Runs: 3 Time: 1:00 Status:
Tumor Sim..
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Task Tumor Sim1
Category Models
Author Yuri Mansury
Date 23 June 2004
Description This is a java module that simulates the growth of a tumor.
Comment In this experiment, we are varying the cell migration rate to approach the affect witnessed by Chaitanya's recent growth exepriment.
Annotations
Collections
Add
Inputs (6)
Cell doubling time
Tumor migratory
Apoptosis rate
Cell metabolism
Glucose coeff
Metabolite coeff
Outputs (3)
sim.out.txt
opendxParameters.dx
Image 2D sequence
Overview Inputs Outputs Workflow

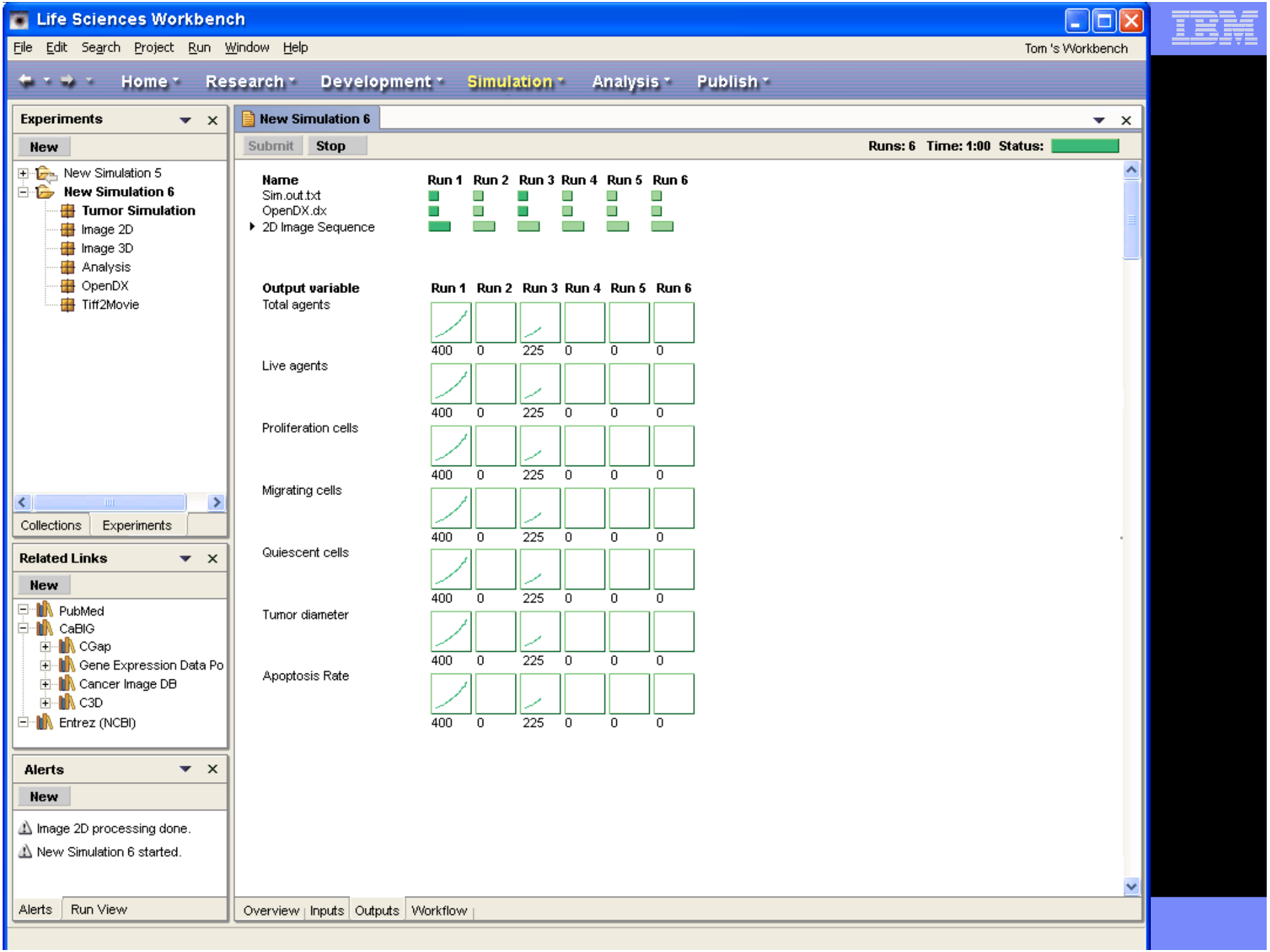

Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 6
Run 1 2 3
Tumor simulation
Image 2D
Image 3D
Alerts Run View
New Simulation 6
Submit Stop
Runs: 6 Time: 1:00 Status:
Parameter Value Units Annotation
Cell doubling time 400.0 xx Cell doubling time 38+-4 hrs
500.0 xx
600.0 xx
Tumor migratory 0.7 xx Tumor cell migratory from Brown
Apoptosis rate 0.01 xx Tumor cell apoptosis rate 23
Cell metabolism 0.1 xx Tumor cell metabolism rate 16
Glucose coeff 0.001 xx New
Metabolite coeff 0.001 xx Toxic metabolite diffusion coeff
Overview Inputs Outputs Workflow


Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 6
Run 1 2 3
Tumor simulation
Image 2D
Image 3D
Alerts Run View
New Simulation 6
Submit Stop
Runs: 6 Time: 1:00 Status:
Tumor Sim..
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Task Tumor Sim1
Category Models
Author Yuri Mansury
Date 23 June 2004
Description This is a java module that simulates the growth of a tumor.
Comment In this experiment, we are varying the cell migration rate to approach the affect witnessed by Chaitanya's recent growth exepriment.
Annotations
Collections
Add
Inputs (6)
Cell doubling time
Tumor migratory
Apoptosis rate
Cell metabolism
Glucose coeff
Metabolite coeff
Outputs (3)
sim.out.txt
opendxParameters.dx
Image 2D sequence
Overview Inputs Outputs Workflow


Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 6
Run 1 2 3
Tumor simulation
Image 2D
Image 3D
Alerts Run View
New Simulation 6
Submit Stop
Runs: 6 Time: 1:00 Status:
Image 2D
Tumor Sim..
Tiff2Movie
Create alert
Simulation New Simulation 6
Module Image 2D
Delivery Alert
IM
Email
Run All
When After
Message Image2D processing done.
Cancel OK
Task Tumor Sim1
Category Models
Author Yuri Mansury
Date 23 June 2004
Description This is a java module that simulates the growth of a tumor.
Comment In this experiment, we are varying the cell migration rate to approach the affect witnessed by Chaitanya's recent growth exepriment.
Annotations
Collections
Add
Inputs (6)
Cell doubling time
Tumor migratory
Apoptosis rate
Cell metabolism
Glucose coeff
Metabolite coeff
Outputs (3)
sim.out.txt
opendxParameters.dx
Image 2D sequence
Overview Inputs Outputs Workflow


Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 6
Run 1 2 3
Tumor simulation
Image 2D
Image 3D
Alerts Run View
New Simulation 6
Submit Stop
Runs: 6 Time: 1:00 Status:
Tumor Sim..
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Task Tumor Sim1
Category Models
Author Yuri Mansury
Date 23 June 2004
Description This is a java module that simulates the growth of a tumor.
Comment In this experiment, we are varying the cell migration rate to approach the affect witnessed by Chaitanya's recent growth exepriment.
Annotations
Collections
Add
Inputs (6)
Cell doubling time
Tumor migratory
Apoptosis rate
Cell metabolism
Glucose coeff
Metabolite coeff
Outputs (3)
sim.out.txt
opendxParameters.dx
Image 2D sequence
Overview Inputs Outputs Workflow


Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Run View
New
New Simulation 6
Run 1 2 3
Tumor simulation
Image 2D
Image 3D
Alerts Run View
New Simulation 6
Submit Stop
Runs: 6 Time: 1:00 Status:
Image 2D
Tumor Sim..
Image 3D
Analysis
Alert
Simulation: New Simulation 6
Module: Image 2D
Run: 3
Image 2D processing done!
Tiff2Movie
Task Tumor Sim1
Category Models
Author Yuri Mansury
Date 23 June 2004
Description This is a java module that simulates the growth of a tumor.
Comment In this experiment, we are varying the cell migration rate to approach the affect witnessed by Chaitanya's recent growth exepriment.
Annotations
Collections
Add
Inputs (6)
Cell doubling time
Tumor migratory
Apoptosis rate
Cell metabolism
Glucose coeff
Metabolite coeff
Outputs (3)
sim.out.txt
opendxParameters.dx
Image 2D sequence
Overview Inputs Outputs Workflow


Life Sciences Workbench
File Edit Search Project Run Window Help
Tom 's Workbench
Home Research Development Simulation Analysis Publish
Experiments
New
New Simulation 5
New Simulation 6
Tumor Simulation
Image 2D
Image 3D
Analysis
OpenDX
Tiff2Movie
Collections Experiments
Related Links
New
PubMed
CaBIG
CGap
Gene Expression Data Po
Cancer Image DB
C3D
Entrez (NCBI)
Alerts
New
Image 2D processing done.
New Simulation 6 started.
Alerts Run View
New Simulation 6
Submit Stop
Runs: 6 Time: 1:00 Status:
Name Run 1 Run 2 Run 3 Run 4 Run 5 Run 6
Sim.out.txt
OpenDX.dx
2D Image Sequence
Output variable Run 1 Run 2 Run 3 Run 4 Run 5 Run 6
Total agents
400 0 225 0 0 0
Live agents
400 0 225 0 0 0
Proliferation cells
400 0 225 0 0 0
Migrating cells
400 0 225 0 0 0
Quiescent cells
400 0 225 0 0 0
Tumor diameter
400 0 225 0 0 0
Apoptosis Rate
400 0 225 0 0 0
Overview Inputs Outputs Workflow