# NanoInformatics Knowledge Commons Curation Team

*Building the Data Dictionary to Enrich Data Curation*

CEINT
Center for the Environmental Implications of NanoTechnology

***Yuan Tian***

***Jaleesia Amos***

Duke UNIVERSITY

# Presentation Contents

***NIKC Defined Curation:*** How does telling the transformational story of a nanomaterial aid data analysis?

***The Data Dictionary:*** How do we leverage collaborations to develop a process for building the data dictionary?

***Visualization Tool:*** Demonstration of the NIKC data dictionary visualization tool.

***Advancing Data Curation:*** How can we use the visualization tool to make data curation easier?

# CEINT NIKC Defined Curation

NIKC-based curation involves:

- Organized storing of data using informed-decision making to determine:
  - *What information* should be curated
  - When data curation is *complete*
  - How to identify *high-quality* data

Purpose of NIKC structure & curation process:

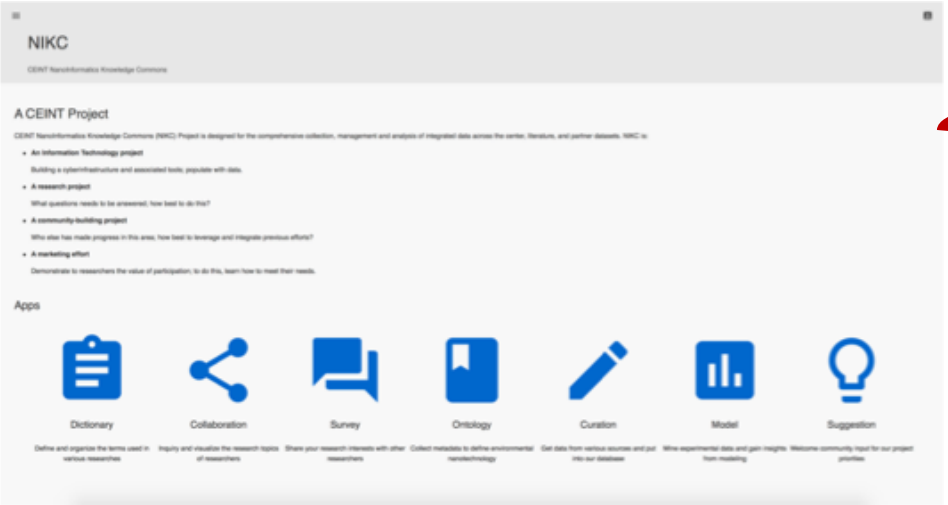
- Tell the story of nanomaterial transformations
- Queryable datasets



Hendren, C.O., Beilstein J. Nanotechnol. 2015, 6, 1752–1762.

# NIKC Instance Organizational Structure Purpose

**NanoInformatics Knowledge Commons Database**



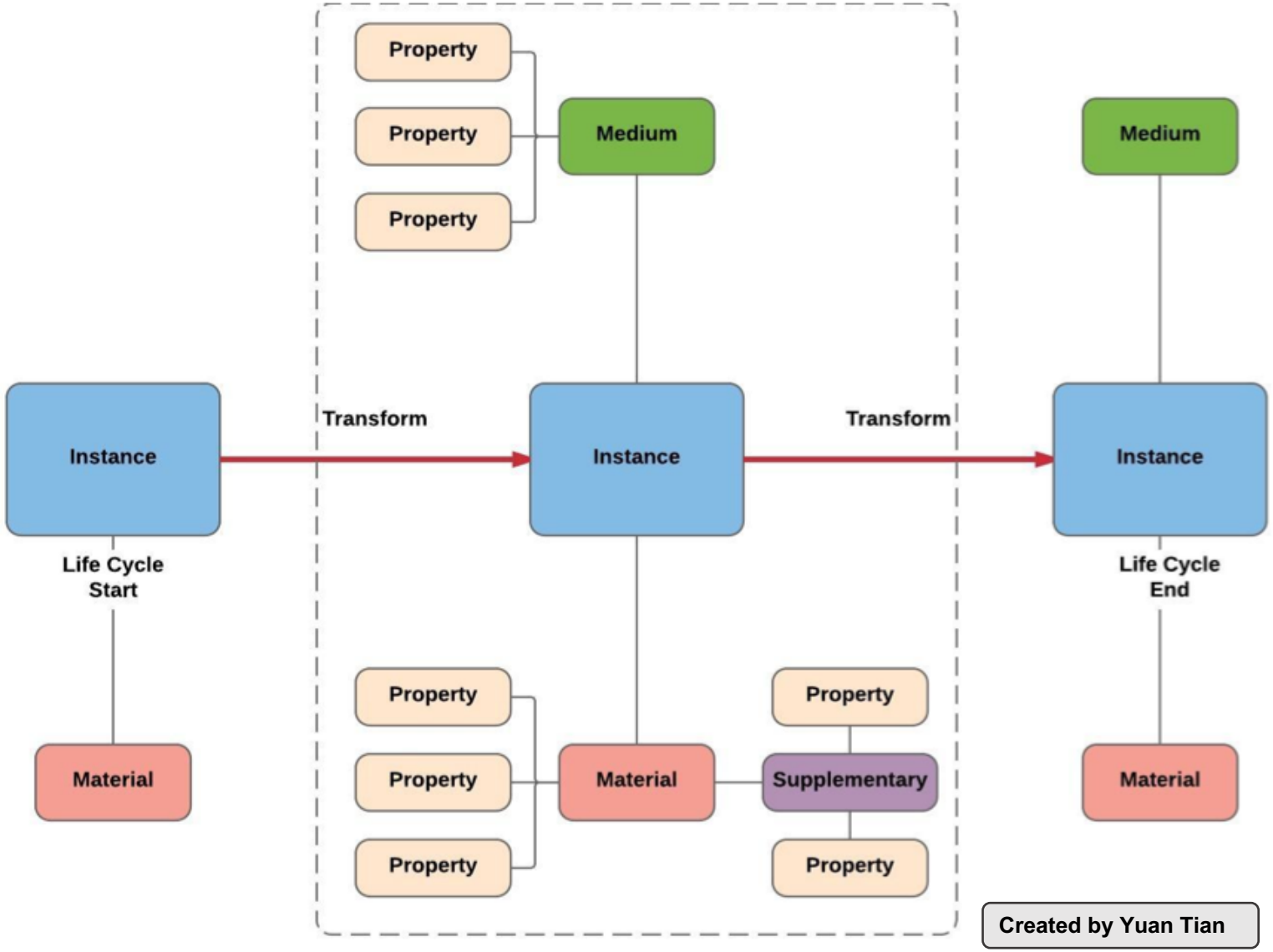

**Nano Product Hazard and Exposure Assessment Tool**


- To elucidate the general principles that determine ***nanomaterial behavior*** in the environment
- To identify data and metadata necessary to support ***forecasts of exposure potential, bioaccumulation, and bioactivity***
- To identify key ***measurement assays*** that are predictive of outcomes of interest

# NIKC Curation: Categorizing Data

| | |
|---|---|
| **Instance** | Describes the chemical or biological properties of the nanomaterial and the medium at a specific moment in time. |
| **Material** | The nanomaterial being studied. |
| **Medium** | Describes the nanomaterial's environment(s). |
| **Supplementary** | Digital images, documents, or tables. |
| **Property** | Can be used to describe the instance, medium, material, or supplementary. |

# NIKC Instance Organizational Structure



# Instance Map: A Transformational Story

Created by Zhao Zhang



NA Monteiro-Riviere et al. Toxicol Lett 155 (3), 377-384. 2005 Mar 15.

# Relating NIKC IOS to Literature Curation


Toxicology and Applied Pharmacology 235 (2009) 216–225

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

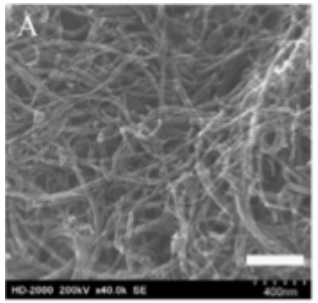
journal homepage: www.elsevier.com/locate/ytaap

Acute and long-term effects after single loading of functionalized multi-walled carbon nanotubes into zebrafish (*Danio rerio*)

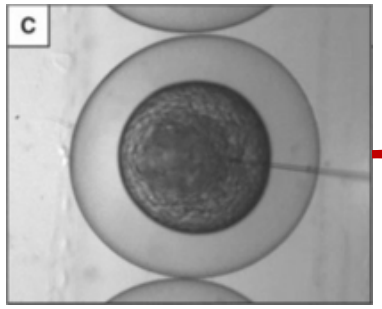
Jinping Cheng [a], Chung Man Chan [a], L. Monica Veca [d], Wing Lin Poon [b], Po Kwok Chan [c], Liangwei Qu [d], Ya-Ping Sun [d,*], Shuk Han Cheng [a,*]

[a] Department of Biology and Chemistry, City University of Hong Kong, Hong Kong
[b] Department of Biochemistry (Science), Chinese University of Hong Kong, Hong Kong
[c] Department of Knowledge Transfer Office, City University of Hong Kong, Hong Kong
[d] Department of Chemistry and Laboratory for Emerging Materials and Technology, Clemson University, Clemson, South Carolina 29634-0973, USA

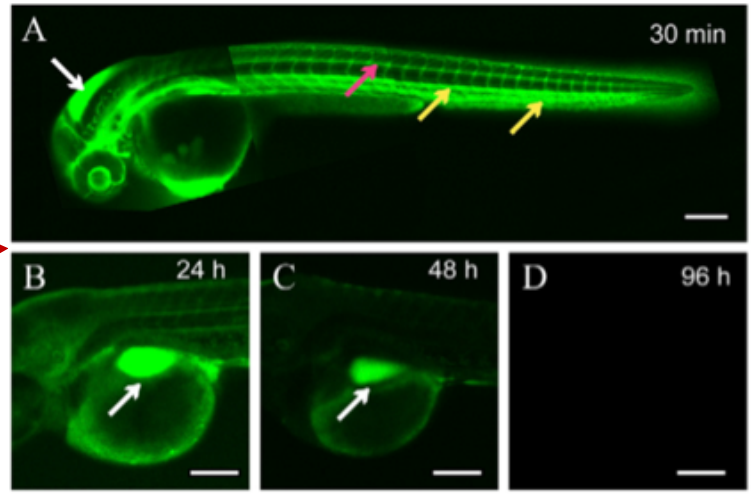
FITC-BSA-MWCNT



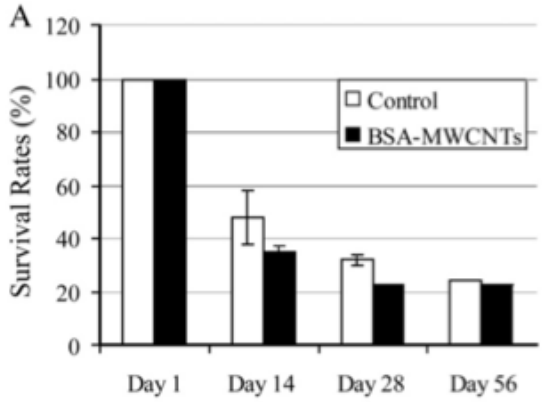

Exposure




Biodistribution




Survival

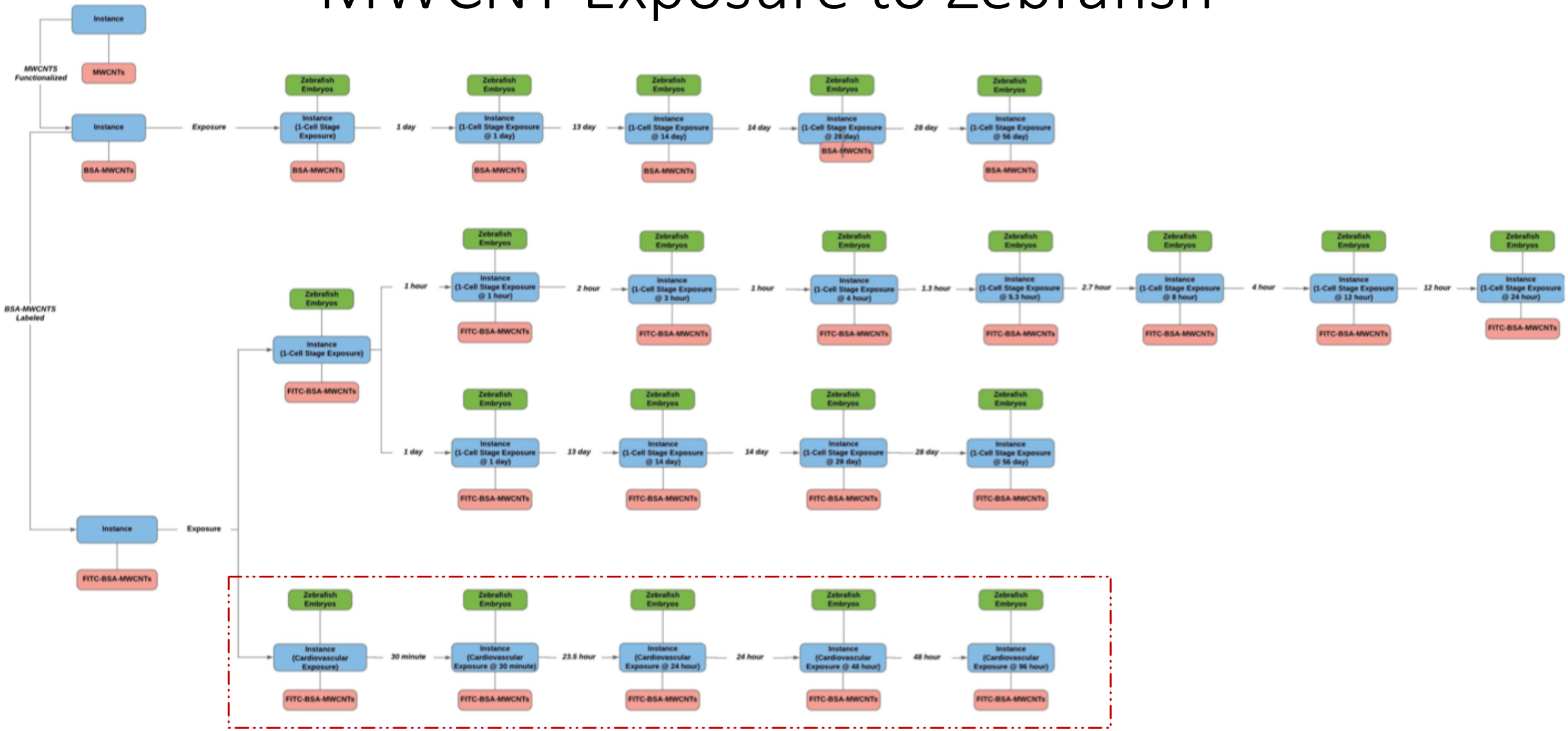





1. Cheng, Jinping, et al. Toxicology and Applied Pharmacology 235 (2), 216-225. 2009 Mar 1.

2. Hogan B.M., et al. (2008) Manipulation of Gene Expression During Zebrafish Embryonic Development Using Transient Approaches. In: Vincan E. (eds) Wnt Signaling. Methods in Molecular Biology, vol 469. Humana Press, Totowa, NJ

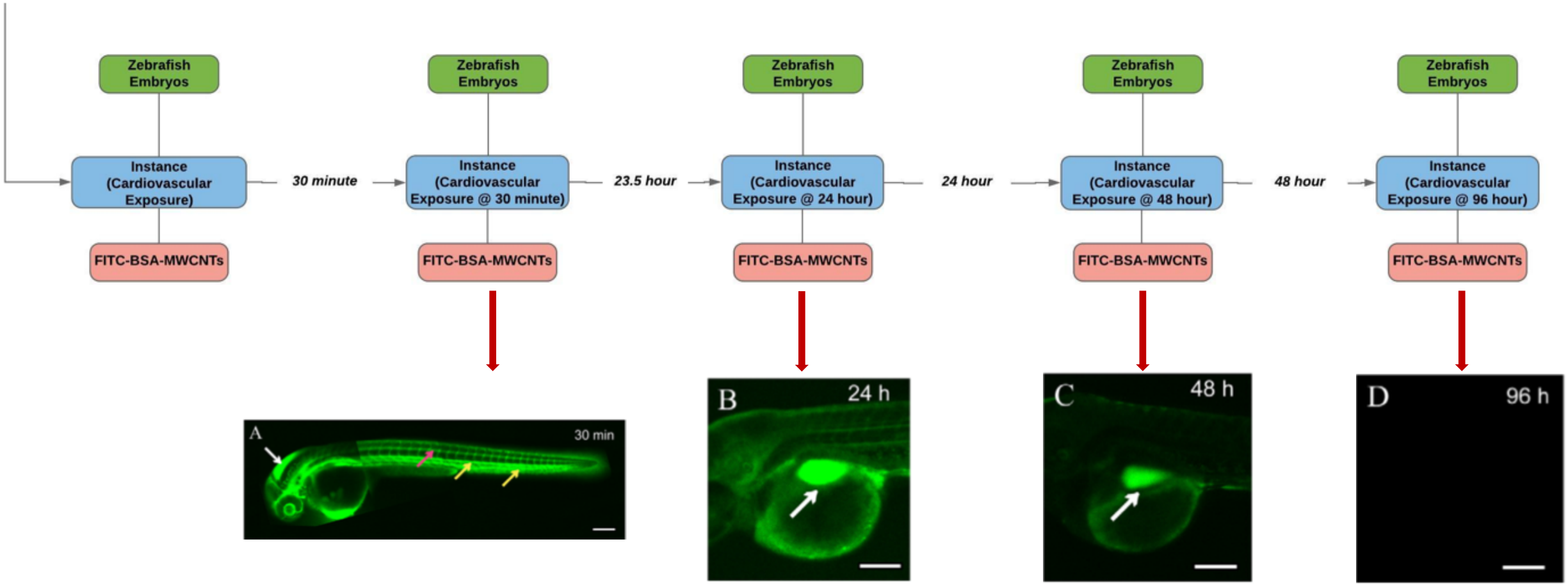
# MWCNT Exposure to Zebrafish






1. Cheng, Jinping, et al. Toxicology and Applied Pharmacology 235 (2), 216-225. 2009 Mar 1.

# Track Measurements Over Multiple Timepoints






1. Cheng, Jinping, et al. Toxicology and Applied Pharmacology 235 (2), 216-225. 2009 Mar 1.

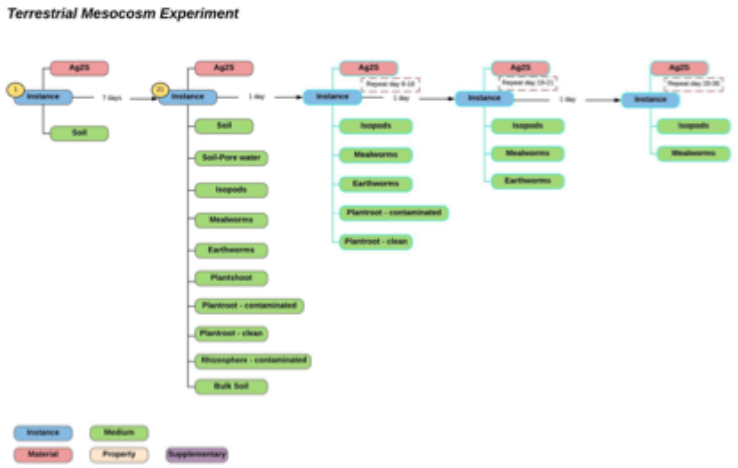
# Relating NIKC IOS to Ongoing Research

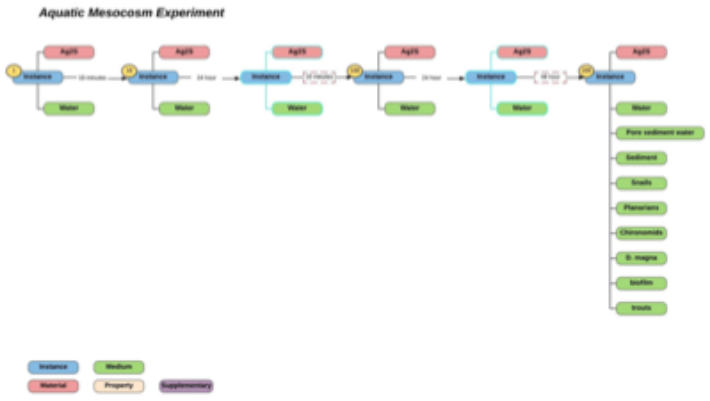
## Mesocosm Template

| datasetId | measurementId | measurementType | referencingId | timeRelative | timeUnit | timeAbsolute | parameter |
|---|---|---|---|---|---|---|---|
| | 90 | property | 43 | | | | total zinc |
| | 91 | property | 43 | | | | total zinc |
| | 92 | property | 43 | | | | total copper |
| | 93 | property | 43 | | | | total copper |
| | 94 | property | 43 | | | | total copper |
| | 95 | property | 43 | | | | total copper |
| | 96 | property | 43 | | | | total iron |
| | 97 | property | 43 | | | | total iron |
| | 98 | property | 43 | | | | total iron |
| | 99 | property | 43 | | | | total iron |
| | 100 | instance | 16 | | | 2018-04-24T 09:00:00 | |
| | 101 | material | 100 | | | | core component |
| | 102 | property | 101 | 10 | minute | | total upper Ag concentrati |
| | 103 | property | 101 | 10 | minute | | total upper Ag concentrati |
| | 104 | property | 101 | 10 | minute | | total upper Ag concentrati |
| | 105 | property | 101 | 10 | minute | | total upper Ag concentrati |
| | 106 | property | 101 | 24 | hour | | total upper Ag concentrati |
| | 107 | property | 101 | 24 | hour | | total upper Ag concentrati |
| | 108 | property | 101 | 24 | hour | | total upper Ag concentrati |
| | 109 | property | 101 | 24 | hour | | total upper Ag concentrati |
| | 110 | property | 101 | 10 | minute | | total middle Ag concentrat |
| | 111 | property | 101 | 10 | minute | | total middle Ag concentrat |
| | 112 | property | 101 | 10 | minute | | total middle Ag concentrat |
| | 113 | property | 101 | 10 | minute | | total middle Ag concentrat |

Bibliography | Publication | Institution | People | Measurement | Protocol | Instrument | description | dictionary

## Instance Map




NanoFASE

Created by Tassos Papadiamantis

# Differences in Curation Methods

## *Literature Curation*

**Identify Quality Data**
*CEINT interests*

**Create Instance Map**
*Use as a curation guide*

**Extract Data**
*Graphs, Tables, Images*

**Enter Data into Excel Template**

## *Ongoing Research*

**Collaborate with Researchers**
*Identify key measurements*

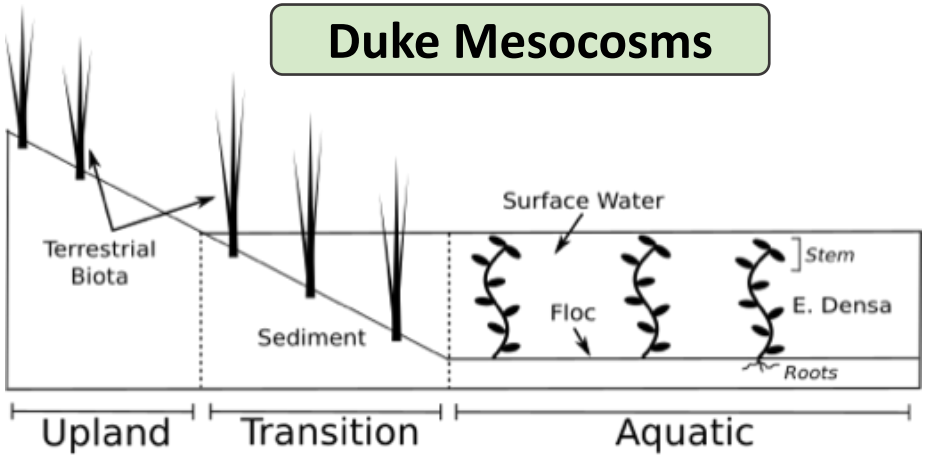
**Create Experiment Template**

**Create Instance Map**
*Verify template Structure*

**Give Template to Researchers**
*Enter Data into Excel Template*

# NIKC IOS Structure Advantage

**Duke Mesocosms**




Shared by Nick Geitner


**Experimental Questions**

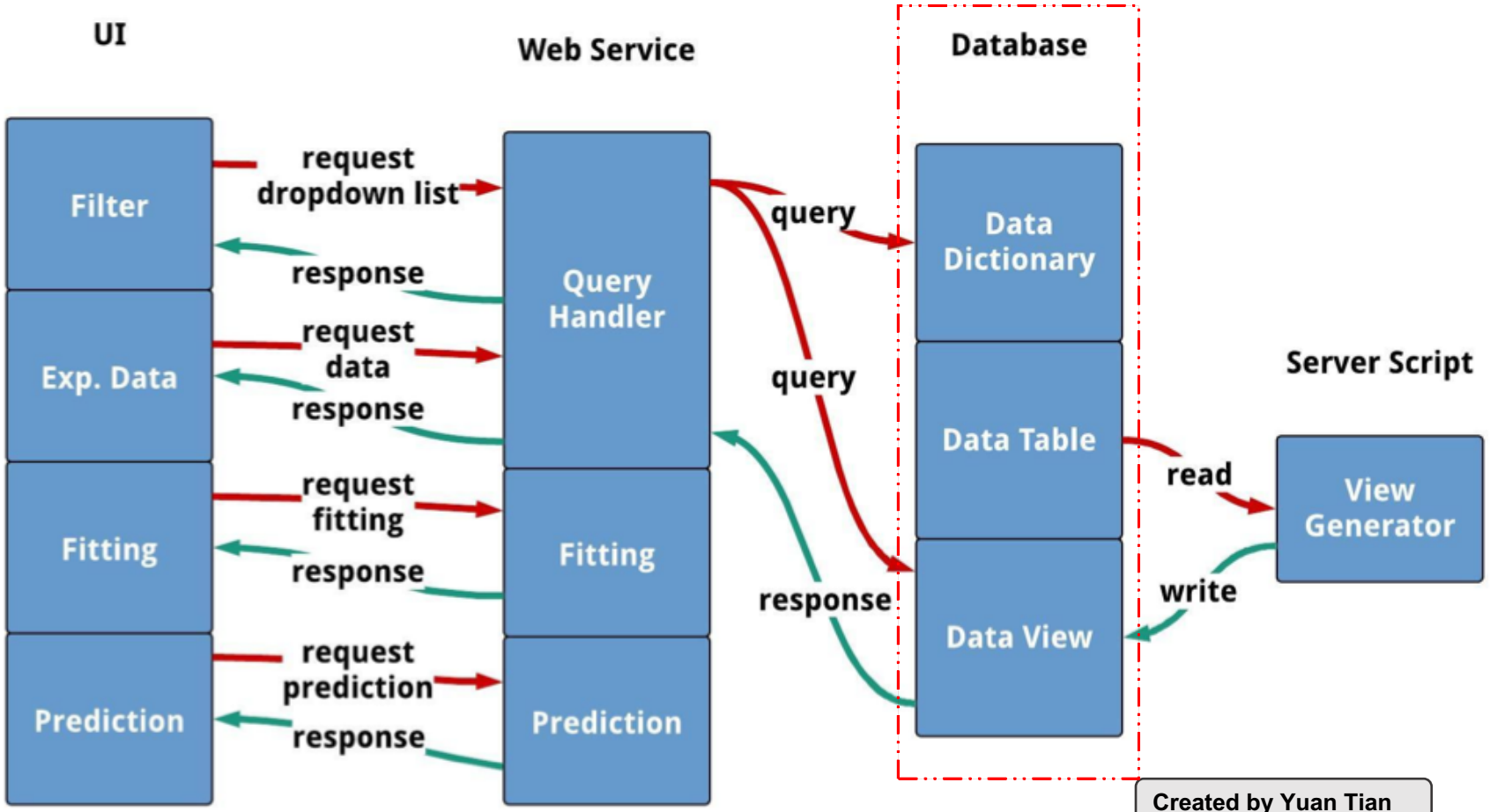
- *How does the media properties affect the nanomaterial's fate?*
- *What factors contribute most to speciation changes?*
- *Does introduction location of the nanomaterial change distribution?*

**Risk & Hazard**

- *How do we design safer nanomaterials?*
- *How do we regulate nanomaterials to reduce exposure risks?*

# CEINT NIKC Web Application Workflow

Web Application includes multiple components running on the web browser, database, and web server.




Created by Yuan Tian

# Approaches to Building a Data Dictionary

***Necessary for Curation***

- Consistency in parameter term names
- Queryable by the database and NanoPHEAT
- Essential for interoperability between databases
- Identify non-ontology terms

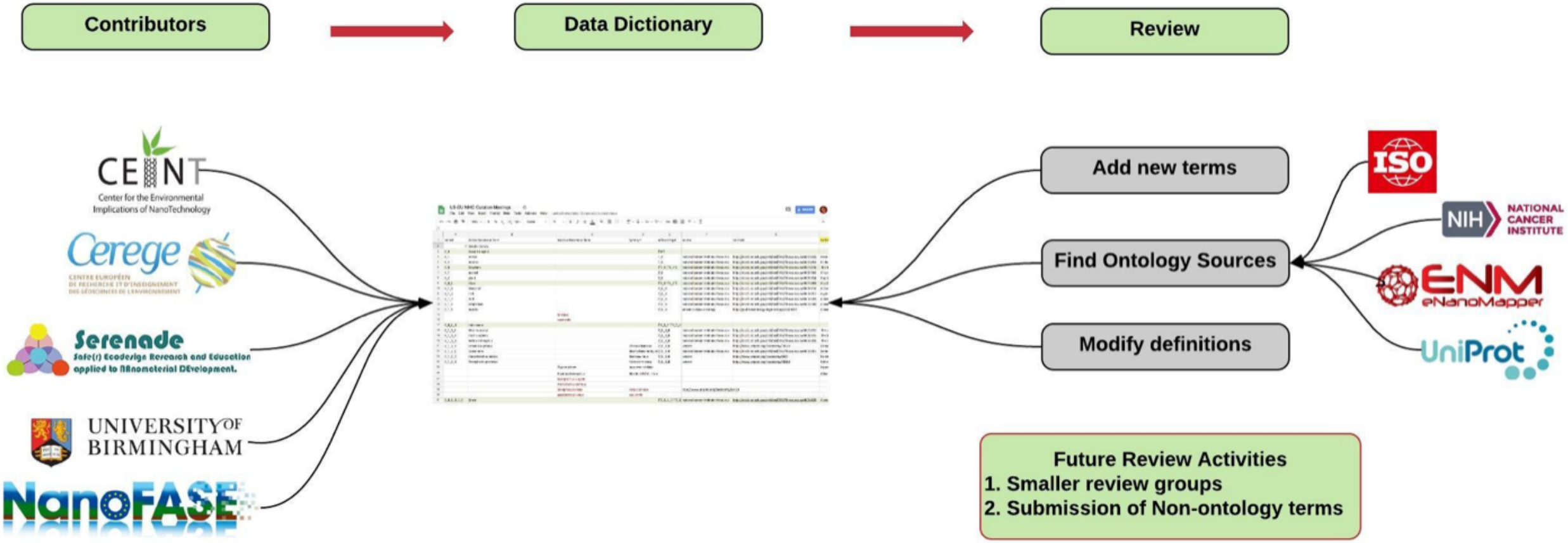
**Build Data Dictionary**
***Start with curated terms***

**Build Hierarchy Structure**
***Ontologies, Expert Elicitation***

**Develop Policies for Use**
***Inclusion and Exclusion of terms***

**Monthly Maintenance**
***Continually expanding***

# Partnered Growth of Data Dictionary



# Future Development of the Data Dictionary

## *Expansion of Data Dictionary*

- Regulatory terms
- Protocols and instruments
- Exposure
- Medium

**POC: Tassos Papadiamantis**

## *Dissolution*

***Smaller Review Group***

- Determine an acceptable curation format.
- Identify necessary metadata for dissolution protocols.
- Identify terms that need to be further defined or eliminated.

**Created by Fred Klaessig**


**ISA-TAB-Nano Expanded: Data Submission Templates**

https://ceint.duke.edu/research/nikc/isa-tab-nano

**Alpha & Dissolution Template**
**Led by Nancy Birkner**

# NIKC Data Dictionary Visualization Tool

**Tool Demonstration by Yuan Tian**

### *Data Curation*

Link Data Dictionary to instance map

### *Data Dictionary*

Automate the Data Dictionary review process

### *Interoperability*

Create a shareable format that can be used for future tool development

# Presentation Summary

*CEINT has created the NanoInformatics Knowledge Commons Database and associated analytical tool, which can be leveraged by collaborators for individual research needs.*

*Growing the NIKC's curated datasets and building the Data Dictionary is a fluid process that is occurring in parallel collaborative efforts.*

*Open to further collaborations: US-EU Curation Meetings, Thursdays 9:00 EST.*

# Acknowledgements


**CEINT NIKC Team**

Yuan Tian
Lila Thornton
Zhao Zhang
Nancy Birkner
Christine Ogilvie Hendren
Mark Wiesner
Greg Lowry

**UoB & CEH Team**

Tassos Papadiamantis
Iseult Lynch
Claus Svendsen
Eva Valsami-Jones
Marianne Matzke

*The combined effort of over 130 researchers*

**SERENADE Team**

Camille de Garidel-Thoron
Jérôme Rose
Armand Masion
Mélanie Auffan


Thank You for Listening!