

NanoInformatics Knowledge Commons US-EU Data Integration Team

A Collaborative Approach to Building Rich Datasets

Jaleesia Amos



Camille de Garidel-Thoron



Tassos Papadiamantis



UNIVERSITY OF
BIRMINGHAM

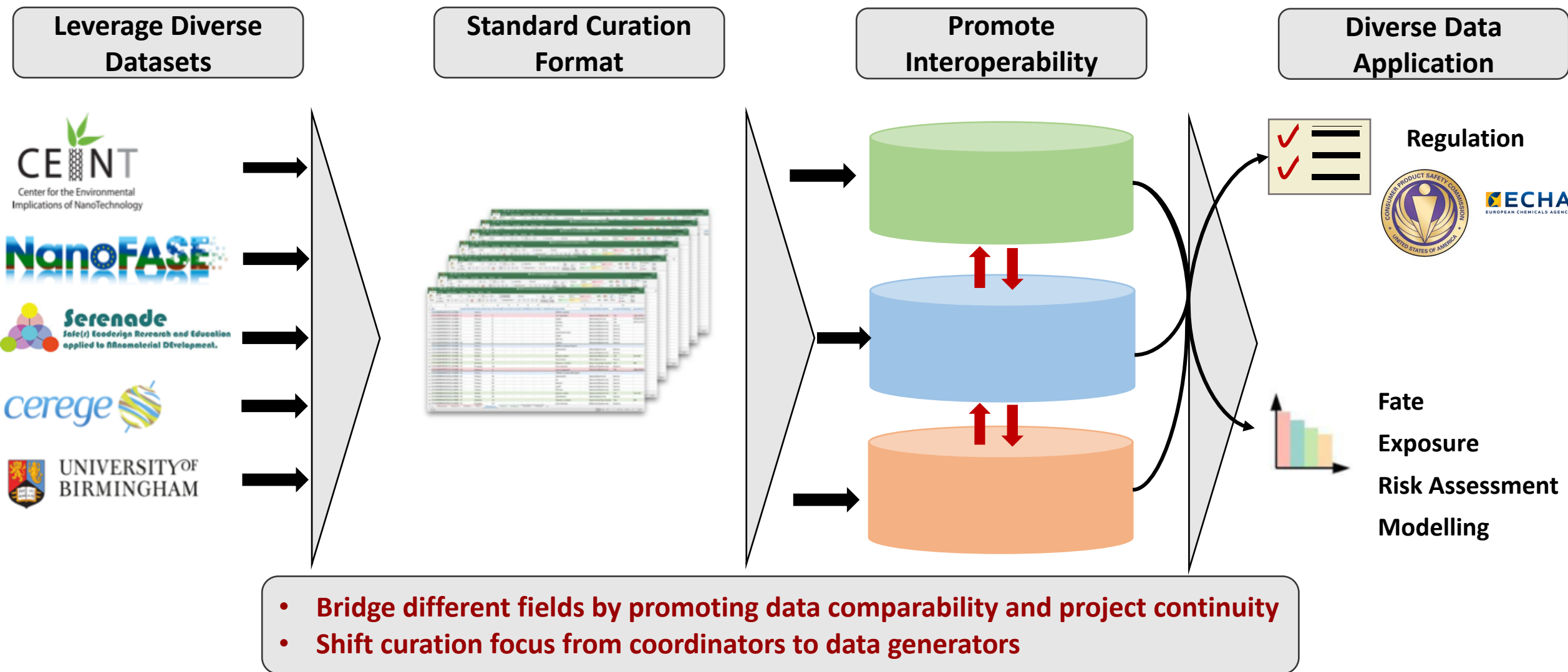
Nano  Group



Data Integration Contents

- Harmonizing Curation Efforts: How do we unite diverse datasets with similar interests to create a set of common goals?
- Data Integration Process: How do we integrate data using curation?
- Advancing Nanoinformatics: How do we move forward by standardizing data curation formats?

Creating Interoperability through Data Curation



Data Integration Process

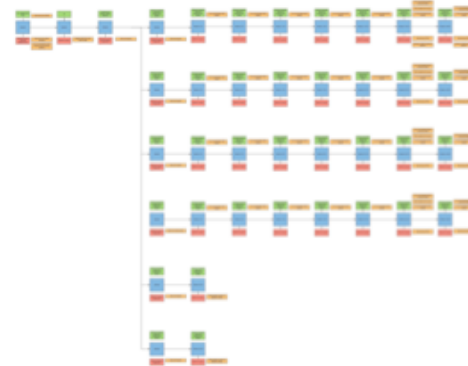
1 Define Goals



2 Choose Format



3 Teach Curation



4 Data Dictionary



5 Communication



How do we define “Link”?

Technical Aspects of Data Linkage

Which members from each project need to meet?

How compatible are the databases currently for linkage?

Exchange template spreadsheets and template manual.

Policy Aspects of Data Linkage

How do we address ownership rights/embargoed publications?

Data Integration Team's Collective Interests

SERENADE • CEREGE

- To develop an **integrated database** and the associated visualization tools
- To allow the researchers to analyse their own data and **integrate them into other dataset**
- To investigate the role of multiple parameters in **predicting behavior** and fate of NMs and their potential risk
- Consumer and environmental **exposure datasets** (aging of products experiments, mesocosms)

CEINT

- 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

NanoFASE • NanoCommons

- To promote **data harmonisation**, continuity and **comparability**
- Create harmonised curation **templates** containing UoB's ENM library, to be used by projects
- Use the curated data to uncover underlying patterns and test data **translation across species**
- Use **data translation** to create larger datasets, which are more robust to the requirements of statistical tests

Diverse Datasets

SERENADE • CEREGE

Material: - Inorganic NP
(Metal, Metal oxides,
Nanoclay, Imogolite)

Type: - Safe by design, **Exposure**,
End of life, **Toxicity, Risk
Assessment**

Studies: - in vitro, in vivo, aging
experiments, mesocosms,

CEINT

Material: - Nanosilver
- Carbon Nanotubes

Type: - **Exposure, Fate,**
Toxicity, Biouptake,
Functional Assays

Studies: - *In vitro*
- *In vivo*

NanoFASE • NanoCommons

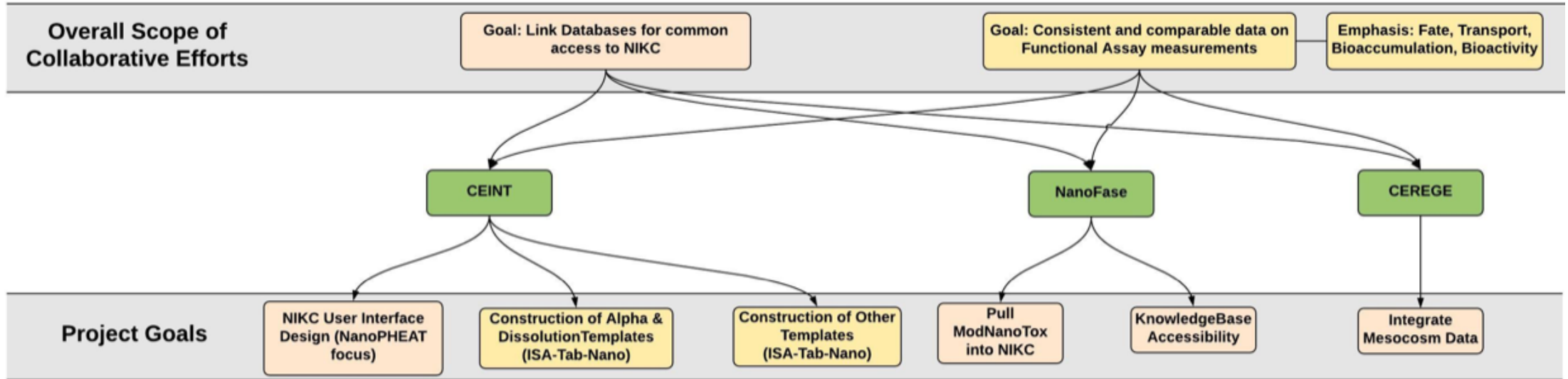
Material: Ag, metal oxides,
chemically doped ENM, ENM
mixtures, hydroxylapatite

Type: Physicochemical,
structural, computational
characterization; **exposure; fate;**
ageing; risk assessment;
toxicity; transformation

Studies: *in vitro, in situ, in vivo*

NanoInformatics
Knowledge Commons
Database

Our Initial Goals



Data Integration Process

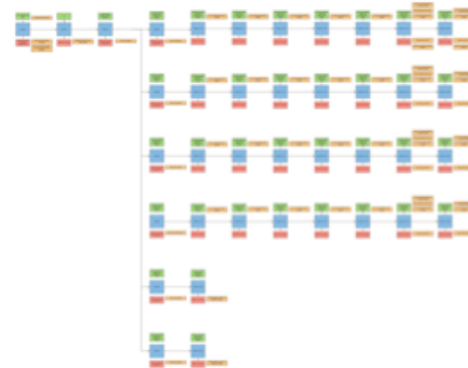
1 Define Goals



2 Choose Format



3 Teach Curation



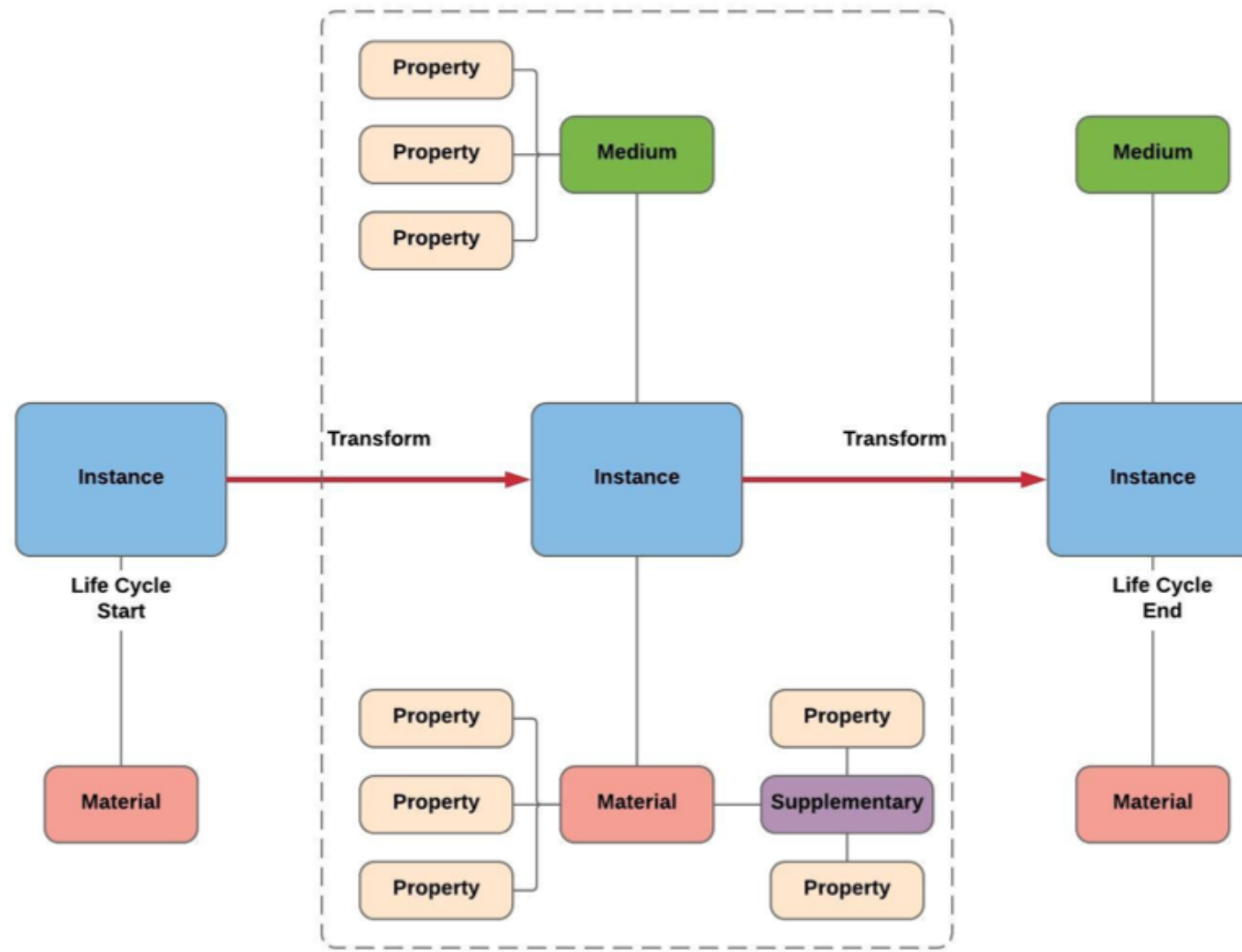
4 Data Dictionary



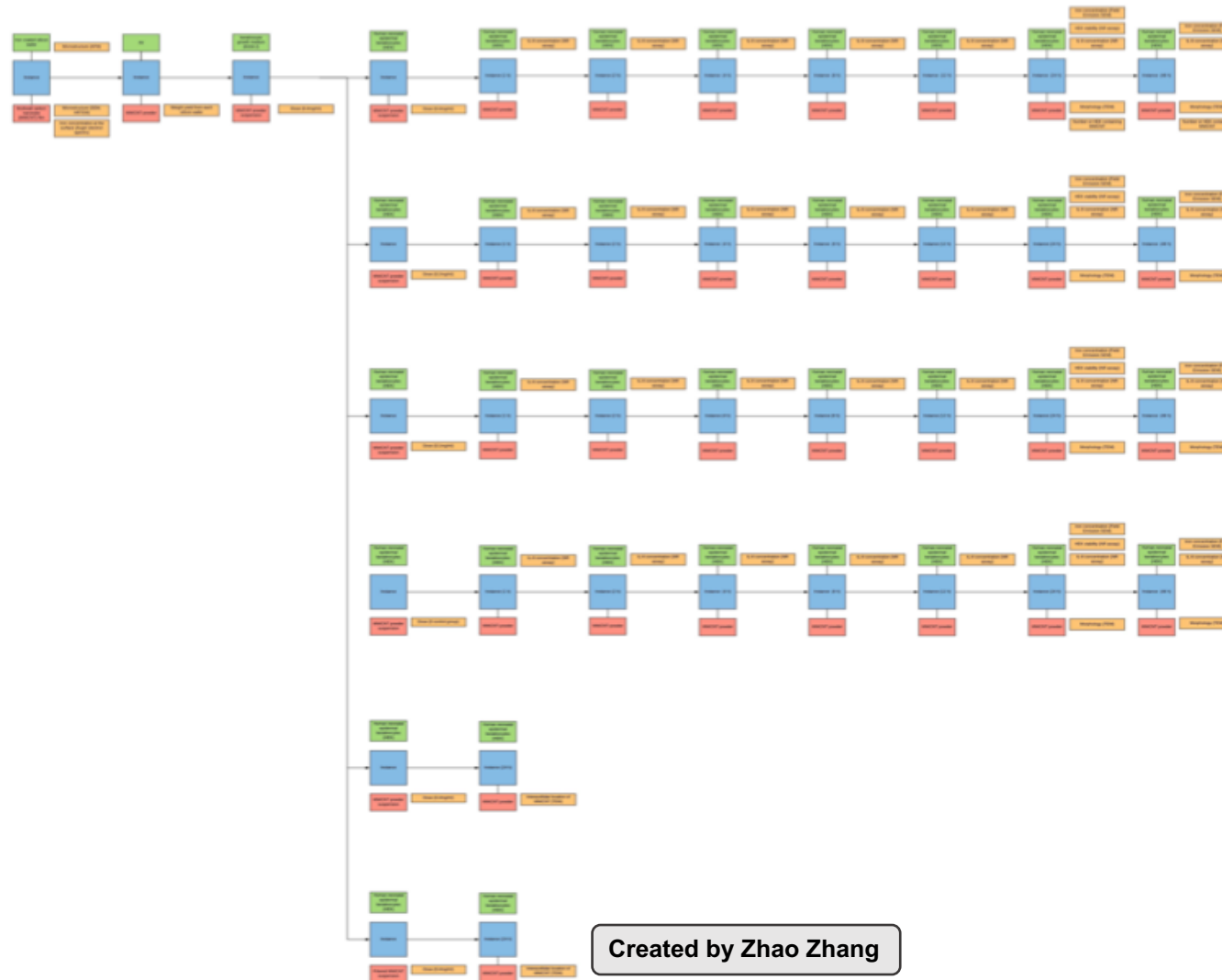
5 Communication



NIKC Instance Organization Structure



Instance Map: A Transformational Story



Data Integration Process

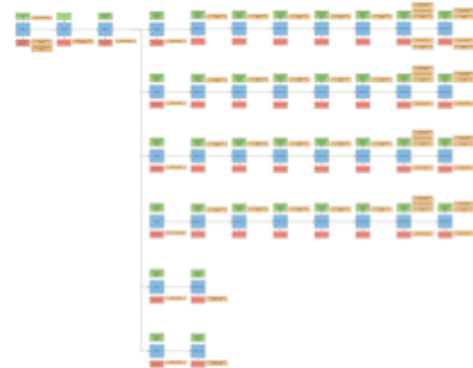
1 Define Goals



2 Choose Format



3 Teach Curation



4 Data Dictionary



5 Communication

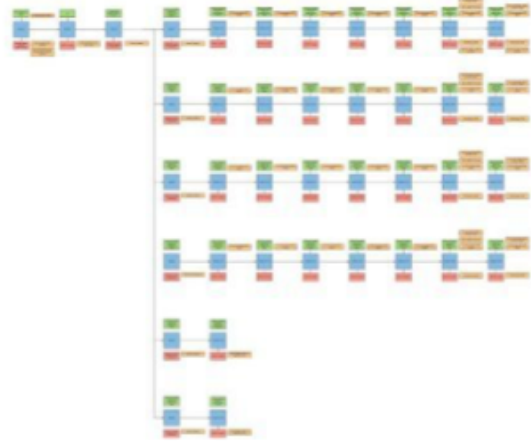


NIKC Defined Data Curation

Collect Data

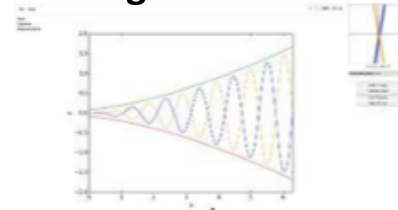


Create Instance Map



Fill out NIKC Template

Digitize Data



Data Dictionary



Time: ~ 1 hour

Time: ~ 2 hours

Time: ~ 2 days

Methods for Teaching Data Curation

Side-by-Side Training

Time: ~ 2 days

Session 1

- Conceptual understanding
- How to make an Instance map
- Example paper

Session 2

- Make Instance Map with individual's data
- Translate Map to Template
- Check Results

Relate NIKC IOS to Template

Make Instance Map

Translate Map to Template

Practice Paper

Check Results

Training by Teleconference

Time: ~ 1 week

Session 1

- Conceptual understanding
- How to make an Instance map
- Example paper

Individually

- Make Instance Map with individual's data
- Translate Map to Template

Session 2

- Check Results

Teaching Curation to Data Generators

Side-by-Side Training

- Nanoinformatics practicals : data integration and curation (10 to 20 students by group of 2)
- Hands on training
- Elaboration of the Instance Map
- Data integration into the NIKC excel template
- Checking of the results



- Side by Side training to promote data curation harmonization
- Online workshops for the SERENADE consortium to promote data curation techniques and implanted tools
- Optimization of the curation templates
- Gathering all datasets produced by the Serenade projects (>150 scientific papers)
- Interoperability and links with other databases (especially for environmental exposure datasets)
- Feed in/ developing models to predict behavior and fate of NMs and potential risks for the environment and human health.
- Testing of the robustness of the database
- Communication and support for the users

Data Integration Process

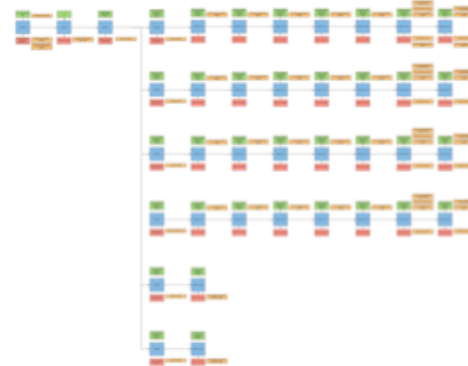
1 Define Goals



2 Choose Format



3 Teach Curation



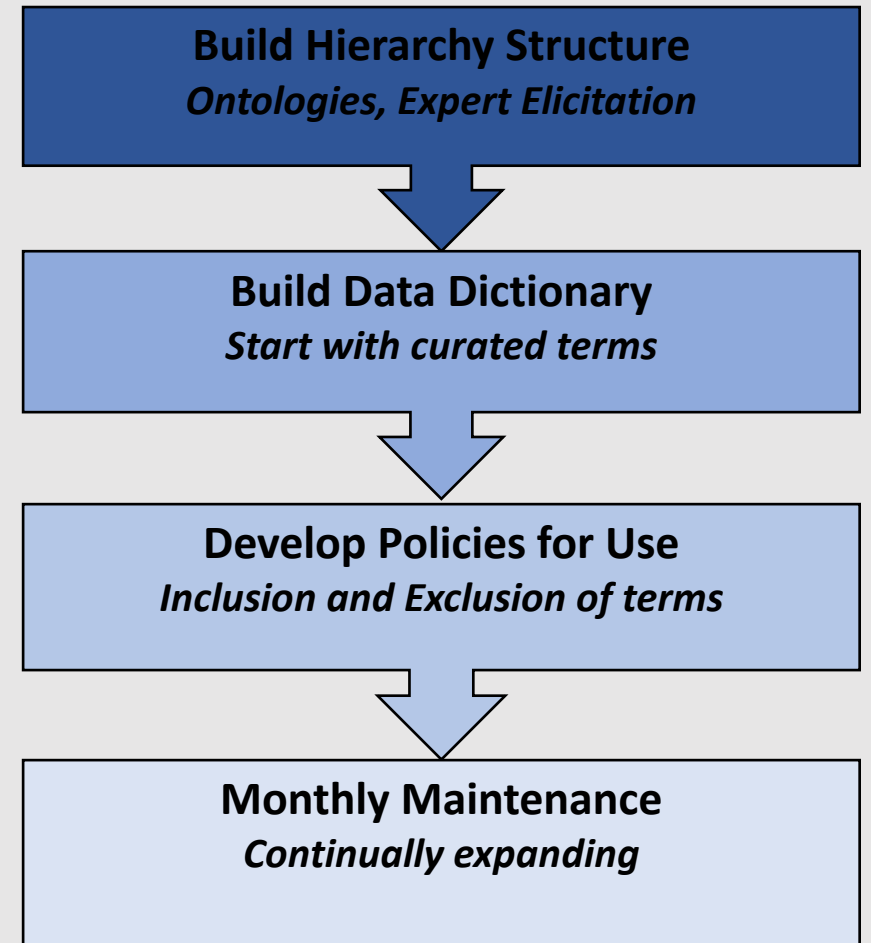
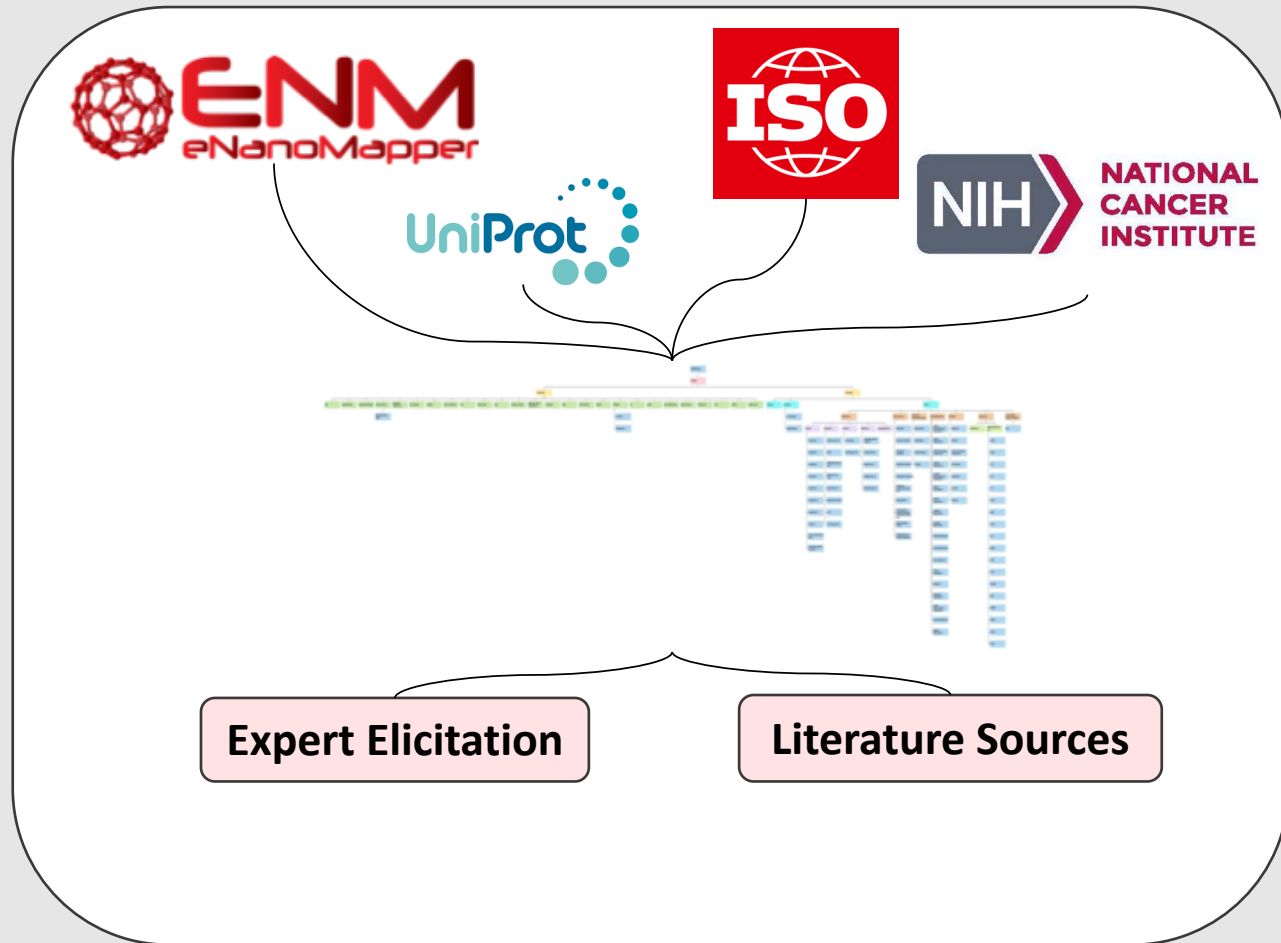
4 Data Dictionary



5 Communication



Approaches to Building a Data Dictionary



Data Integration Process

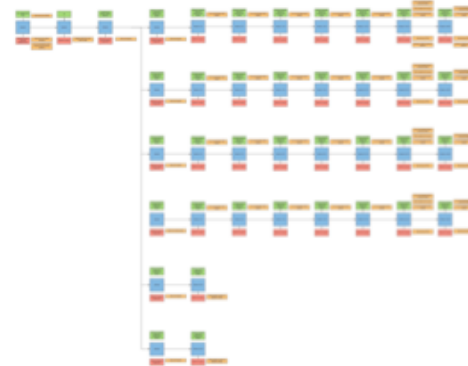
1 Define Goals



2 Choose Format



3 Teach Curation



4 Data Dictionary



5 Communication

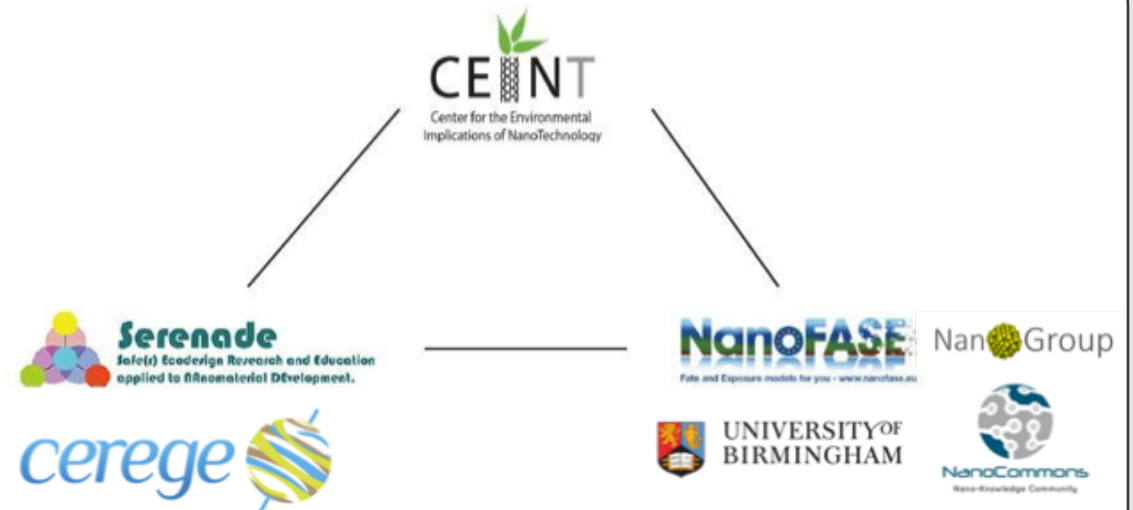


US-EU Curation Meetings

Teleconference

*Thursdays at 9:00 EST/ 2:00 UK/
3:00 FR*

- Discuss curation formatting modifications
- Updates on individual curation efforts
- Updates to data dictionary
- Possible additional collaborations



UoB Approach to Curation



- Promote data curation harmonisation across projects and stakeholders
- Prepare “dynamic” curation templates, which will link to common starting materials and will be EUON compatible
- Create experimental workflows (online- lab books) that will use pre-prepared curation templates
- Automated uploading of curated datasets and implementation of automated analysis tools

- Preparation of curation templates using the characterisation of the NanoGroup ENM library as starting points
- Liaise with members of the group and develop the characterisation templates to fit experimental needs
- Implement an online lab book system using the curation templates for on-the-fly curation and uploading
- Use the curated data to test cross-material and cross-species translational research

- Preparation of curation templates to fit all aspects of the NanoFASE framework for ENM release pathways, fate and transformation, exposure, environmental and bio- nano interactions and modelling
- Implement curated datasets to NanoFASE knowledge base
- Analyse the data to identify patterns and underlying correlations between the various compartments
- Contribute valuable insights for risk assessment and regulatory purposes

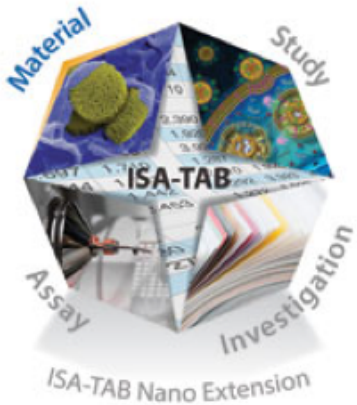
UoB Approach to Curation



- In house pilot testing and optimisation of online lab books with curation templates, automated logging and uploading and analysis tools
- Preparation of a “tiered” detailed data curation guidance document for EU projects and global stakeholders
- In house workshops to implement the curation methods for all of the NanoGroup members and move the curation process to benchtop users
- Face to face and online workshops for the NanoCommons consortium to promote data curation techniques and implemented tools
- Liaising with the NanoFASE consortium to prepare WP specific templates, followed by online and face to face workshops
- Promotion of the developed curation techniques to interested stakeholders. Offer of implementation assistance and teaching (face to face and/or online)
- Continuous communication and support (helpdesk) with potential external users

Moving Forward: Shift Curation Focus onto Data Generators

ISA-TAB-Nano Expanded: Data Submission Templates



Alpha & Dissolution Template
Led by Nancy Birkner

Dissolution_gdoc_FINAL.xlsx
File Edit View Insert Format Data Tools Add-ons Help Last edit was made on September 26, 2017 by Jaleesia Amos

100% - \$ % .0_ .00 123 - Calibri - 11 - B I U A -

Nanomaterial Crystalline Phase Identification (Core)

	CR	CS	CT	CU	CV	CW	CX	CY	CZ	DA	DB	DC	DD	DE	DF	DG	DH	DI	DJ	
1	Nanomaterial sample drying method in preparation for experiment, if appropriate															Experiment Parameters				
2	Was the nanomaterial dried prior to the experiment? (Yes or No)	Nanomaterial Drying Method Reference	Rate of Centrifugation	Units	Centrifuge Temperature	Units	Duration of Centrifugation	Units	Drying Method	Drying Temperature, Minimum	Units	Drying Temperature, Maximum	Units	Drying Time	Units	Nanomaterial Instance	Experiment Protocol (e.g. doi)	Exposure Medium Composition	Exposure Medium Component 1	Exposure Medium Component 2
3																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				

Model for NIKC Data Submission Interface based on Alpha and Dissolution Templates developed by nanocommunity. Outlines reaction systems, media characterizations, nanomaterial transformations, instrument protocols

User Interface Development

Instance 1

NM received as Powder

Instance 1

NM received as Powder

Ag NP

Level 1

Add New Parameter
Describe the nanomaterial and medium of the system

Level 2

Enter Text (e.g. 1-Cell Stage Zebrafish Embryo
Exposure to FITC-BSA-MWCNTs at 1 day)

Level 1

Add New Parameter
Describe any changes to the nanomaterial or medium of the system
What nanomaterial was studied

Level 2

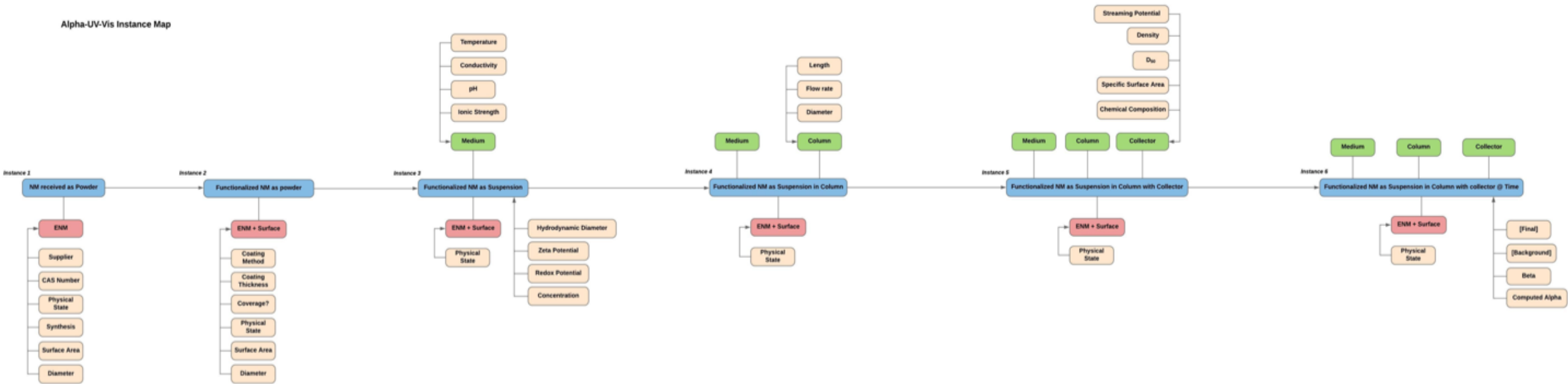
Composition
Core Component
Surface Component

Level 3

Core Component
Silver Nanoparticle (Ag NP)
Carbon Nanotube (CNT)
Single Wall Carbon Nanotube (SWCNT)
Double Wall Carbon Nanotube (DWCNT)
Multi Wall Carbon Nanotube (MWCNT)

User Interface Development: Alpha Column Study

Alpha-UV-Vis Instance Map



Summary: Standardizing Curation for Interoperability

- *How we are using data curation to promote interoperability between projects to diversify data application.*
- *Our process for integrating diverse datasets using one curation format.*
- *Methods we are developing to increase data volume by turning benchtop data generators into curators of their own work.*
- *Using curation as a focal point for interoperability requires working on large and small scales simultaneously.*

Acknowledgements

CEINT NIKC Team

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

UoB & CEH Team

Iseult Lynch
Claus Svendsen
Eva Valsami-Jones
Marianne Matzke

*The combined effort of
over 130 researchers*



SERENADE Team

Jérôme Rose
Armand Masion
Mélanie Auffan



Thank You for Listening!