Global Harmonization of Nanoinformatics:
A Case Study in Convergence and Team Science

Christine Ogilvie Hendren, Duke University
Fred Klaessig, Pennsylvania Bio Nano Systems

NNCO 2019 NanoEHS Webinar Series

November 12th 2019
12-1pm EDT

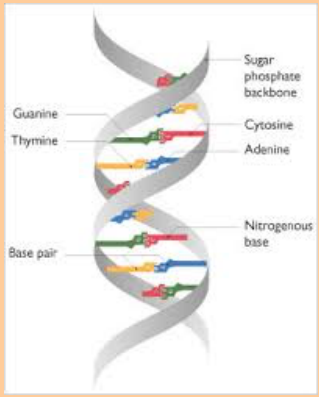
National Cancer Informatics Program

us-eu
bridging nanoEHS research efforts

Too early for big data, but critical time for harmonization to generate comparable datasets going forward.

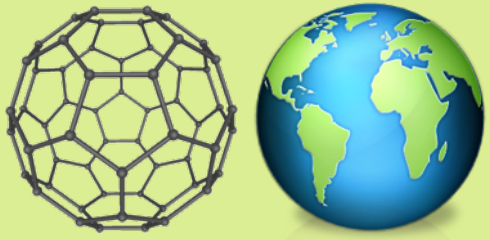
Big Data

- Limited number of established parameters
- Established protocols
- Well understood meta-data
- Huge masses of matching datasets



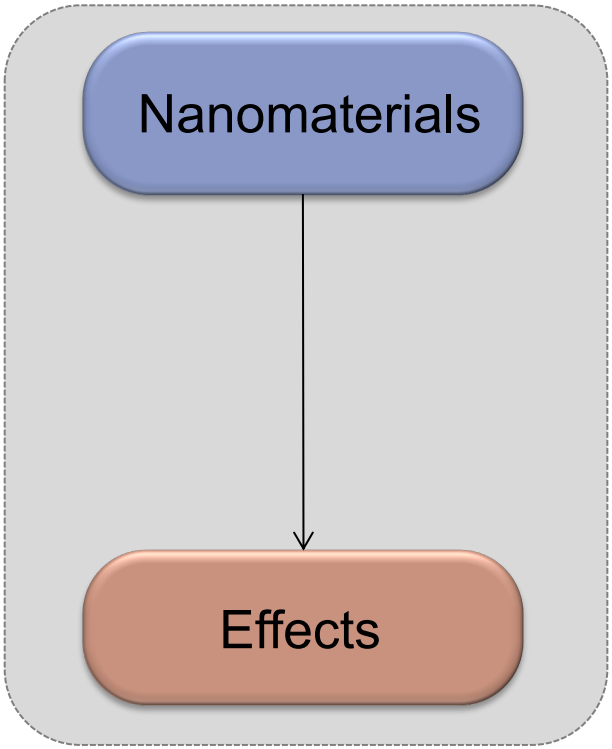
Broad Data

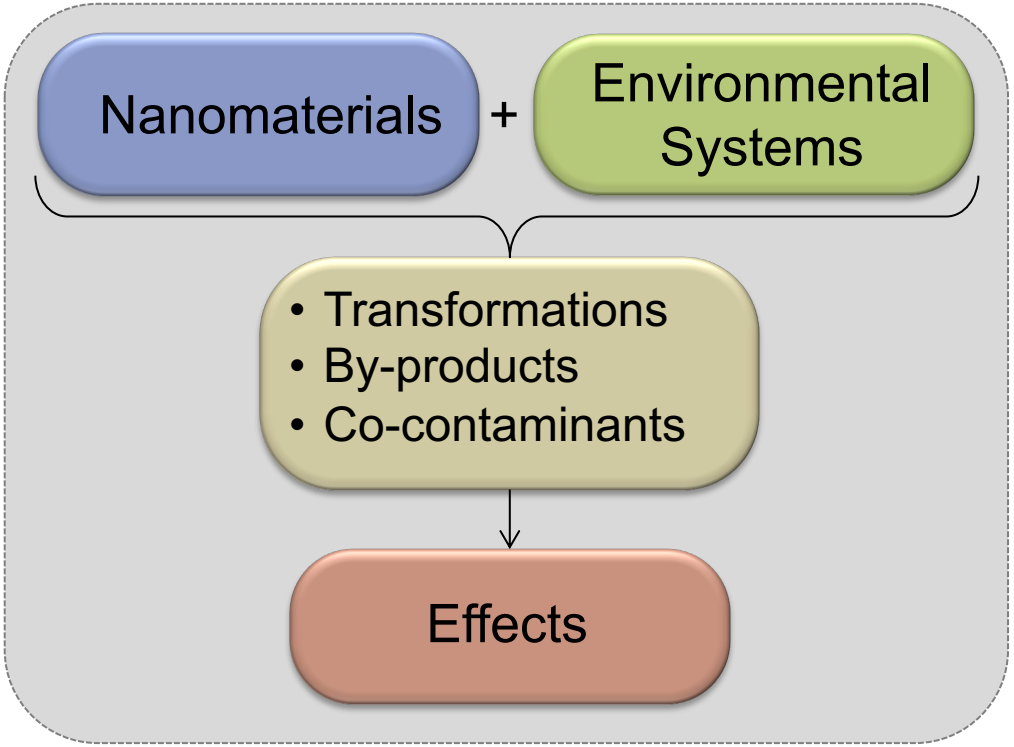
- Huge number of possible parameters, inconsistently measured and reported
- Complex and disparate meta-data
- Masses of mis-matched datasets



Environmental Complexity:

A Necessary Reality to Say Anything Useful About to Nanomaterial Behavior





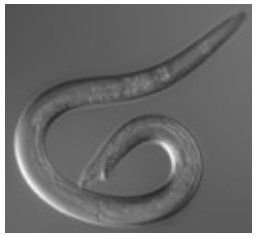
We Must Capture Meta-Data to Understand Nanomaterial Fate, Exposure, Hazard and Risk

e.g. Sulfidation Decreases Toxicity

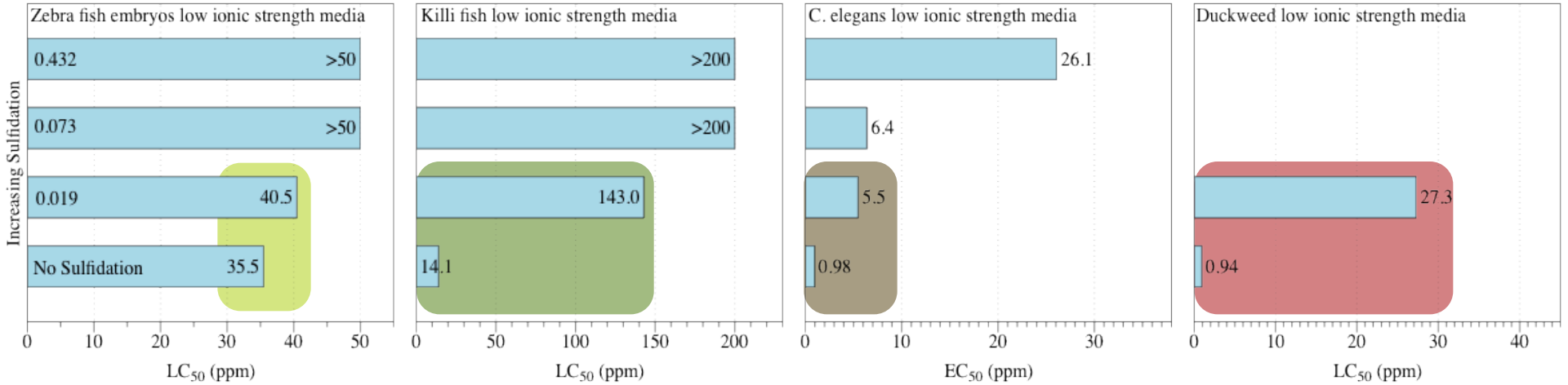
Zebrafish

Killifish

C. Elegans

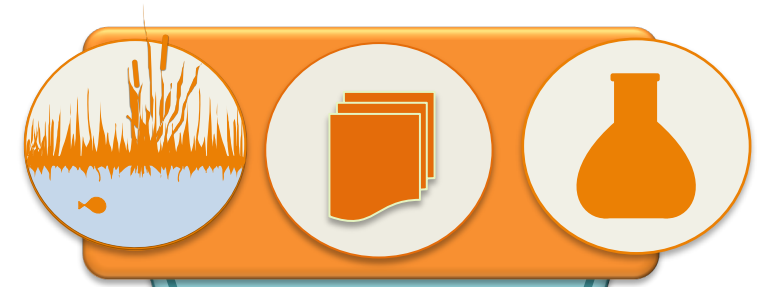


Duckweed



Levard, Hotze, et al., *ES&T 2013, 47, 13440−13448*

Why are we doing any of this?



Collect data

Curators and Researchers Enter Data

…in consistent format
…with consistent language

…to a repository

App Developers & Researchers Analyze

…to query, combine, compare

…to answer questions, and provide guidance

…to pose new research questions

…to combine and compare data across many researchers

Why are we doing any of this?

Collect data

Christine

Curators and Researchers Enter Data

…in consistent format
…with consistent language

…to a repository

…to pose new research questions

…to combine and compare data across many researchers

App Developers & Researchers Analyze

…to query, combine, compare

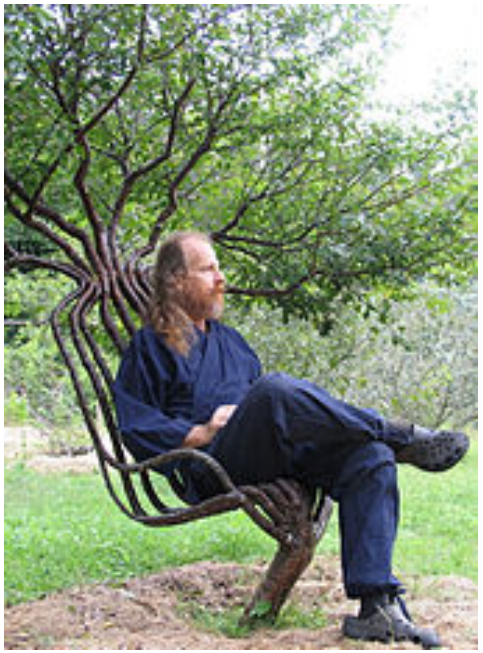
…to answer questions, and provide guidance

Fred

Avoiding Divergence Takes Effort, Consistent Contact, Commitment, and Patience

Divergence in cyberinfrastructure development, ontology, analytical tools, is the norm, and it makes sense.

Each project is separately funded, separately incentivized, and has a finite ”pot” of resources (time + money).



We want the resources to grow together and make each other stronger, leverage each others “nutrients” and becoming magnificent.

If we do it right, we can even use it for things down the road that we can only imagine now.

Many Contributors on the Convergence Journey (Among them, a number of volunteer efforts)

2006
caNanoLab

NanoParticle Ontology, 2010

2014 - present
NIKC

2016 - present
eNM eNanoMapper

2010
NBI KNOWLEDGEBASE

2012
NANOMATERIALREGISTRY

2014 - present
nanoinfo

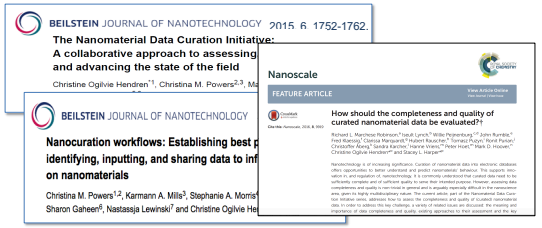
2012
Serenade

2016
NanoFASE

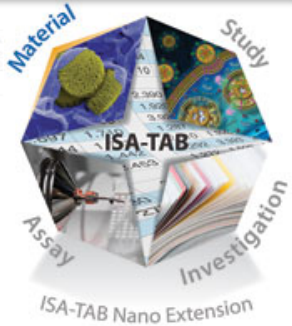
2016
NANOREG

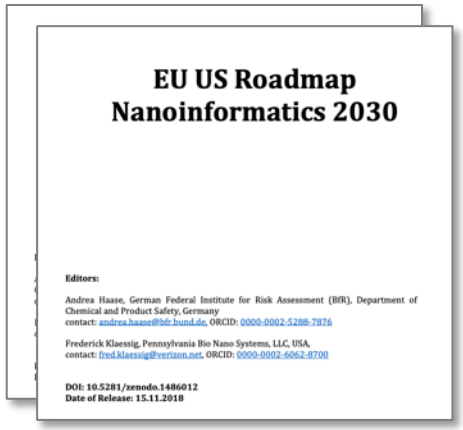
NanoCommons
Nano-Knowledge Community
2018-present

Nanomaterial Data Curation Initiative
2015-2019



ISA-TAB-Nano
2012-13
2016 -17





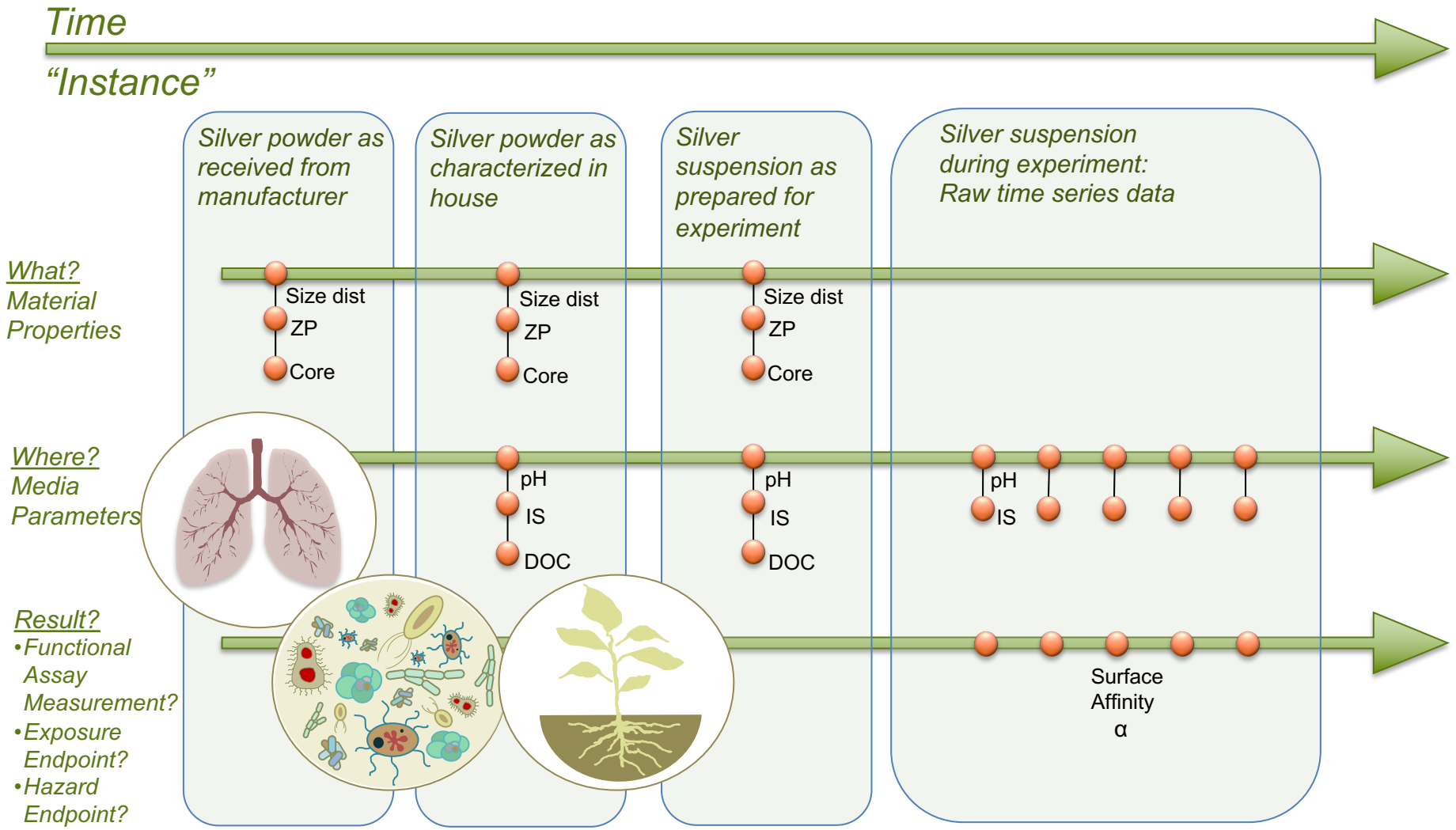
DOI: 10.5281/zenodo.1486012

NCIP
National Cancer Informatics Program
us-eu
bridging nanoEHS research efforts

A[n Abridged] History of Building the NanoInformatics Knowledge Commons

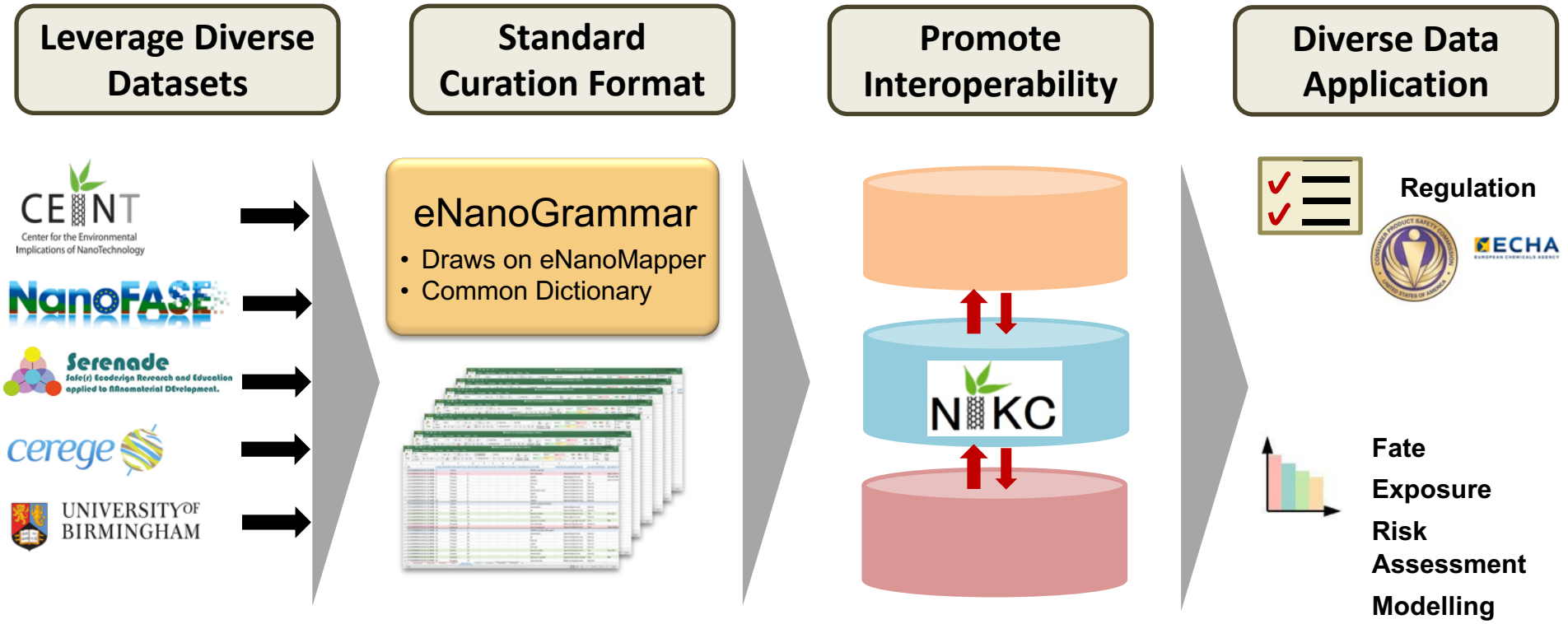
On the NIKC Concept of "Instance"

*Temporally Tracking the Path of a Study**



*Track across stages in a product life cycle and/or across steps within an experiment

Premeditated Interoperability Through Shared Data Curation

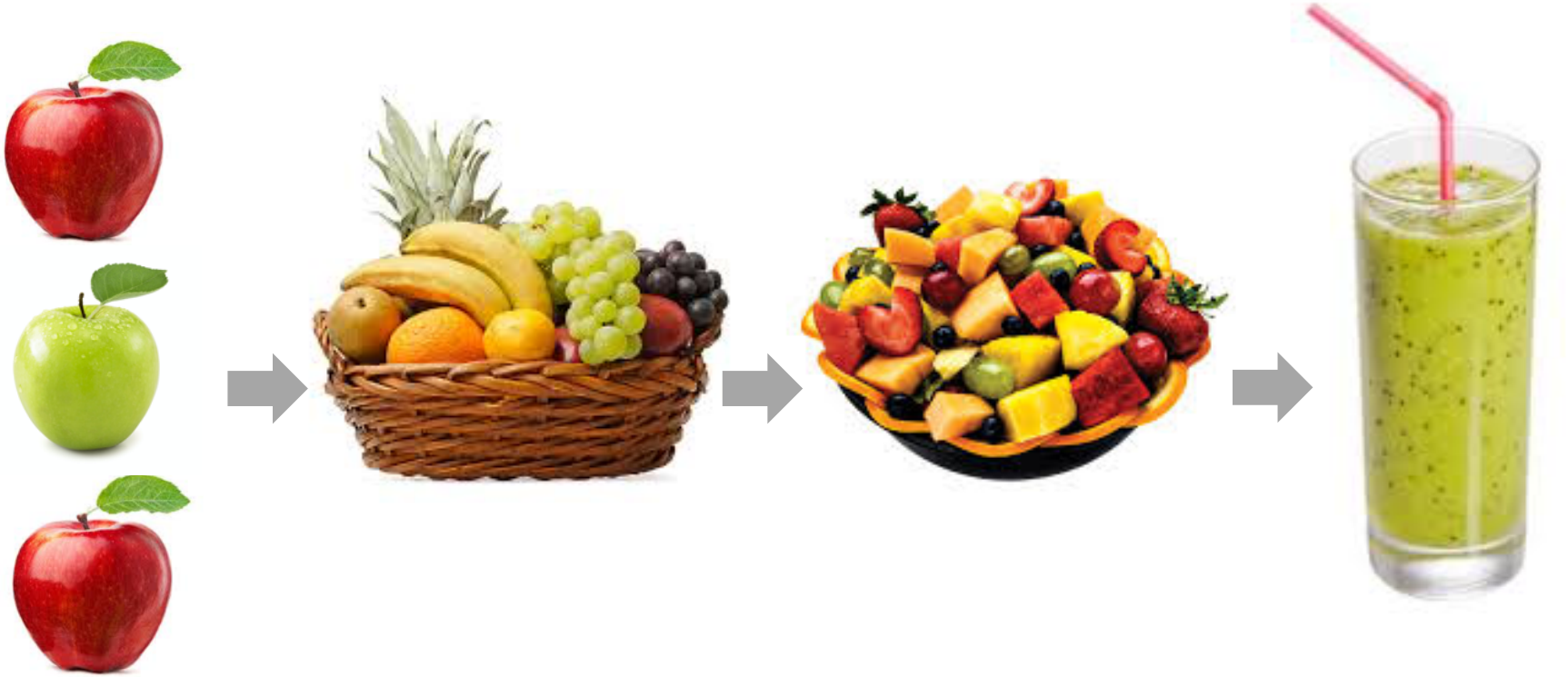


Shared Research Context Across Projects

- Data must support investigation of transformations
- Leverage other work wherever possible
- Modular design to allow targeted app development
- Curation process that allows researchers to focus on research

Convergence:
We need a smoothie for many complex problems

Disciplinary **Multi-Disciplinary** **Inter-Disciplinary** **Trans-Disciplinary**



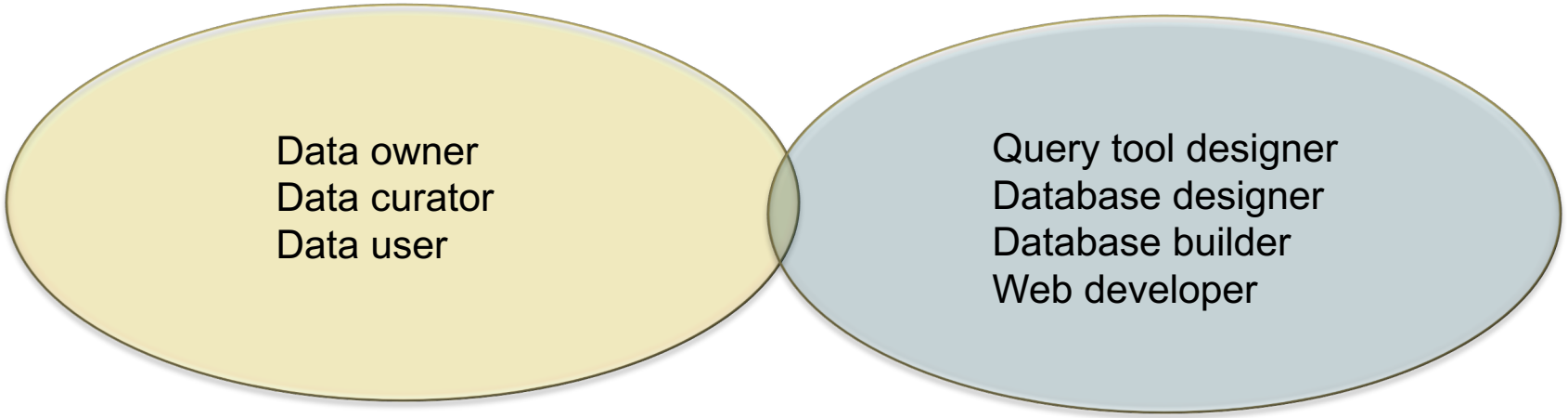
- Convergence is how we get the inextricable mixture needed to address complex problems.
- Diverse teams are necessary for this, and teamwork is a science as well as an art.

Three Keys to Enabling Convergence Exemplified by Nanoinformatics for NanoEHS

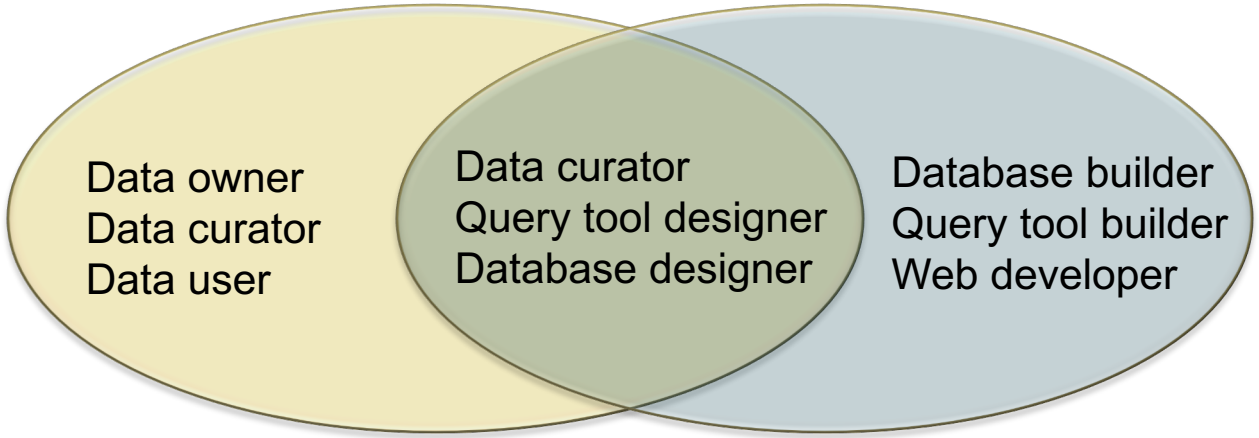
1. Integrate research approaches and methods through **a shared purpose** that promotes harmonization of methods, media, measurements, language
2. Shepherd research-driven **data** integration
3. Invest in Integration and Implementation Science (I2S) **specialists** to coordinate **teams**

But who is the gardener?

With established fields, and due to normal educational pathways, these skill sets are separate.



To work effectively in an emerging science on budgets that also do not include a dedicated and permanent information science staff, this does not work. Science-fluent experts have to do the bulk of the creative and connective work.



Why are we doing any of this?

Collect data

Christine

Curators and Researchers Enter Data

…in consistent format
…with consistent language

…to a repository

…to pose new research questions

…to make sure we get enough of the same data that we can compare

App Developers & Researchers Analyze

…to query, combine, compare

…to answer research questions

Fred

EU-US Roadmap Nanoinformatics 2030

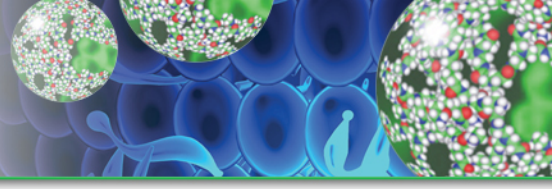
- A survey of database management practices for nanoEHS
  - Editors Fred Klaessig and Andrea Haase
  - 44 contributors; EU, US, Canada, China, Australia
  - Sections 4 & 5: database structure; metadata; data entry issues
  - Sections 6 & 7: data analysis & computational modeling
- Forums:
  - EU-US Community of Research: co-chairs: Klaessig & Willighagen
  - nanoWG: Luisa Sullivan (NIH) & Iseult Lynch (U. Birmingham); Mervi Heiskanen NIH contact (heiskame@MAIL.NIH.GOV)
  - NSC WGF: Egon Willighagen (U of Mastricht)
  - Several papers on work flow; data completeness; etc.

Thursday, 14 Nov., Praktikum

- Hosted by nanoWG
- Egon Willighagen (U of Maastricht) and Nina Jeliazkova (Idea Consult) presenting
- Practical examples
- To join at 11 am EST (17:00 Central European Time)

https://cbiit.webex.com/mw3300/mywebex/default.do?service=1&siteurl=cbiit&nomenu=true&main_url=%2Fmc3300%2Fe.do%3Fsiteurl%3Dcbiit%26AT%3DMI%26EventID%3D202920437%26UID%3D0%26Host%3DQUhTSwAAAARzjk4BlTlu26VwCuJk5JYIeN81IvJ2s2pVA0032Egbfqg952AcE52T7YslTmtUKCve9_D0YXvcKc4dpCCvdni_0%26FrameSet%3D2%26MTID%3Dm91f8beaac1cee4906cd16149e3f6868b

Further Motivation – Nanoinformatics Session, ACS Annual Meeting



Session 1

Talk Title	Presenter
Exploration of the nanomedicine-design space with high-throughput screening and machine learning	Chad Mirkin
Machine-learning driven design of nanomaterials: Ingredients for success	Bryce Meredig
Nanoinformatics in drug delivery: Matching drugs to carriers	Yosi Shamay
caNanoLab: Enhancing retrieval and sharing of cancer nanotechnology data	Luisa Russell
Synthetic closed-loop smart insulin patch	Zhen Gu
Nanoinformatics as a driver for nanoparticle synthesis and biomedical imaging paradigms in MRI and CT	Erik Shapiro

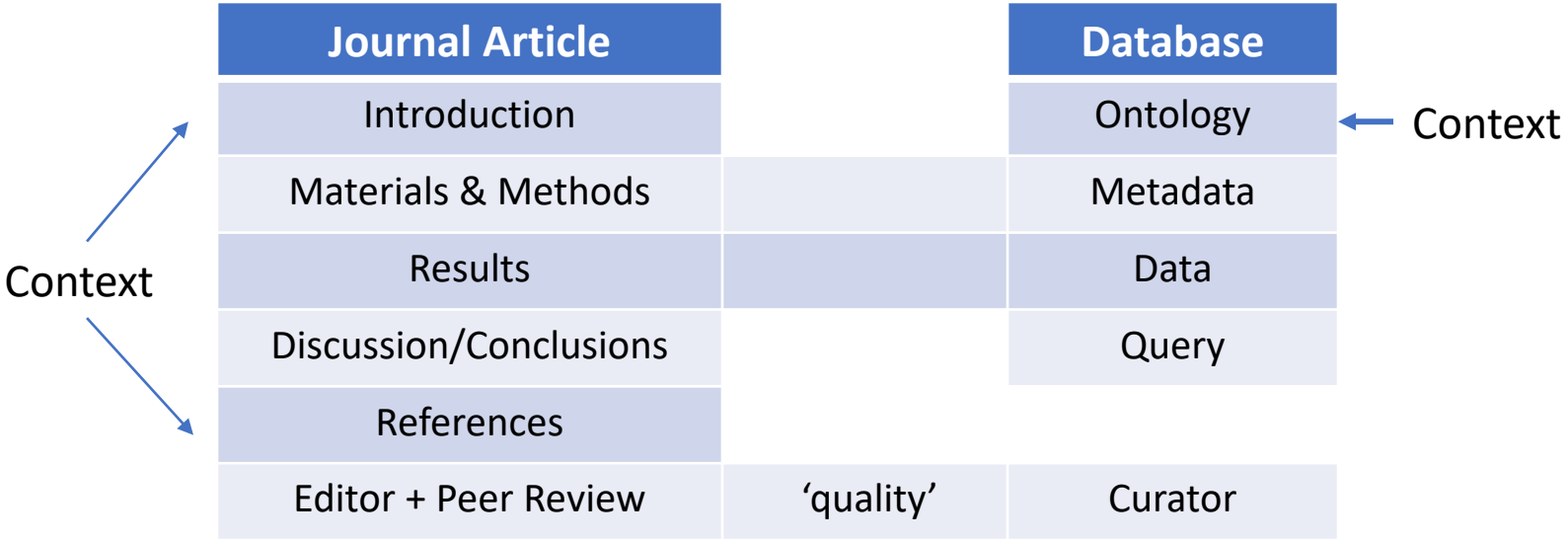
Session 2

Talk Title	Presenter
Experimental and computational search strategies for function in the peptide sequence space	Rein Ulijn
Combinatorial targeting for phenotypic targeting	Giuseppe Battaglia
Transitioning to predictive analysis for nanoparticle biocorona studies	Korin Wheeler
Rapidly identifying nanoparticles for in vivo RNA and gene editing using DNA barcoding	James Dahlman
Learning to predict single-wall carbon nanotube-recognition DNA sequences	Anand Jagota
Chemometric analysis of nanosensors libraries for developing short-wavelength infrared optical probes for anthracyclines	Jackson Del Bonis-O'Donnell
Development of targeted nanomedicines facilitated by nanoinformatics	Daniel Heller

Roadmap Background

1. Significant effort in nanoEHS studies; funding decreasing
2. Many databases expected with different purposes
    1. Federated means there is data exchange using compatible ontology and metadata 'structures'
    2. Specialty: DaNa, providing information on commercial products by application/use; applies regulatory 'filters'
    3. General: nanoinfo.org (UC-CEIN; 600); eNanoMapper (EU Projects; 100); caNano (NIH; 1,000); NIKC (CEINT; 200)
3. Informatics workflow:
    a) Deconstruct laboratory studies…. ***to***
    b) Populate databases…. ***in order to***
    c) Identify patterns & computational models that reconstruct the data for new purposes, i.e. predict toxicity… ***in order to***
    d) Maximize knowledge & limit animal testing

Data that Travel



- With a paper, the author controls context, inferences and later use of inferences (done through monitoring citations).
- With a database, curator controls context & user modulates inferences through queries.
- Reputation (author & Journal) becomes diffuse (anonymous).
- Data-centric research style develops (see Leonelli description)

Roadmap Status & Goals

- Nanoinformatics is one of three roadmaps for funding of EU projects (other roadmaps are EHS & Commercialization).
- Connected to the EU regulatory framework, e.g. portions of eNanoMapper on EUON website administered by ECHA.
- Nanoinformatics Roadmap Goals:
  - Maximize use of nanoEHS data;
  - Catalog ‘best practices’ and challenges across disciplines;
  - Alert the community to regulatory uses of data such as in grouping or read-across; and
  - Provide a coordinated time horizon for regulatory acceptance of nanoEHS data, especially computational models.

Computational Modeling

1. Regulators frequently compare a new chemical substance to analogs with more complete dossiers (QSARs)
2. OECD has criteria for QSAR model acceptability:
    - defined endpoint;
    - unambiguous algorithm;
    - applicability domain;
    - statistically robust; and
    - ‘a mechanistic interpretation, if possible.’
3. Gaining regulatory feedback on models is essential
4. Dissolution of sparingly soluble particles example:
    - Critical for QSARs & PBPK if both particle and dissolution product exhibit adverse effects (dissolution products probably known).

PBTK-Thermo-Dissolution

- PB uses compartments to model physiology
- TK is ADME acting on a toxicant
  - Overall exposure becomes a localized organ dose
  - Can scale across exposures, species, times
  - Mix of kinetic and equilibrium concepts (K_{ow})
- QSPR → QSAR → PBTK →AOPs
- Particles complicate & challenge:
  - Uncertain dose metric & K_{ow} does not apply
  - Dissolution products replenished by solid ($\frac{d\,(c)}{dt} = 0$) and possibly not first order as with molecules
  - Handling adsorption (vascular system & protein corona) open

V_{blood} & Plumbing

$$k_{organ\_up\_ionic} = b_{ionic} * \frac{m_{organ} * c_{organ\_GSH}}{m_{b.w.} * c_{body\_GSH}} \qquad k_{organ\_up\_nano\_cap} = b_{nano\_cap} * \frac{Q_{organ\_blood}}{V_{blood}}$$

- Ionic silver:
  - Eqn. is organ ***mass*** & glutathione concentration
  - The term b_{ionic} is ***[min]$^{-1}$***
  - Calculates a pseudo-partition function for ***body***
- Particulate silver:
  - Equation is ***volume***, not mass, & ***blood***, not body
  - The term b_{nano_cap} is ***dimensionless***
  - V_{blood} is ***series*** pipe configuration, not ***parallel***
  - Allocates by organ blood flow rate, not organ residence time

EFSA Acceptance of GUTS for Plant Protection Products

1. Framework
   - Definitions, equations, ‘accepted’ interpretations
2. Implementation
   - Math package (Mathematica, R)
   - Two ‘ring’ data sets to verify new implementations
3. Selecting case study modules
   - based on experimental design & data
4. Regulator can validate with FOCUS scenarios
   - web accessible Excel implementation from CNRS
5. Epistemic Opacity challenge

Concluding Remarks

1. A functioning Nanoinformatics effort will alter practices on data sharing, analysis and attribution.
2. The curator assumes an important role
3. There will be many types/purposes of databases taking advantage of local maintenance and requiring some degree of federation
4. Consortia should address database issues at the start
5. Regulatory acceptance will be important to industry;
6. Translating ‘research’ data into regulatory formats is recommended for academia

Citations

Slide	Reference(s)
16	• EU US Roadmap Nanoinformatics 2030, Editors: Haase A und Klaessig Fred, DOI:10.5281/zenodo.1486012; https://www.nanosafetycluster.eu/Nanoinformatics2030.html. 2018 (accessed September 2019). • Powers, et al.; *Beilstein J. Nanotechnol.* **2015,** *6,* 1860–1871. doi:10.3762/bjnano.6.189 • Hendren, et al.; *Beilstein J. Nanotechnol*. **2015**, 6, 1752–1762. doi:10.3762/bjnano.6.179 • Marchese Robinson et al.; *Nanoscale*, **2016**, 8, 9919. doi: 10.1039/C5NR08944A • Karcher et al.; *NanoImpact* 9 (2018): 85-101. doi: 10.1016/j.impact.2017.11.002
20	• Leonelli, S.; Data-Centric Biology: A Philosophical Study
21	• Stone et al., 2017; Research priorities relevant to development or updating of nano-relevant regulations and guidelines • Falk et al., 2016, Research roadmap for nanosafety Part III: Closer to the market (CTTM)
22	• OECD PRINCIPLES FOR THE VALIDATION, FOR REGULATORY PURPOSES, OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP MODELS; https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf
23	• Bachler, Gerald, Natalie von Goetz Konrad Hungerbühler, 'A physiologically based pharmacokinetic model for ionic silver and silver nanoparticles.' International Journal of Nanomedicine 2013, 8: 3365–3382
24	• EFSA Journal 2018; 16(8):5377

Acknowledgments

The combined efforts of hundreds researchers

Jaleesia Amos
Nathan Bossa
Nick Geitner
Greg Lowry
Joana Sipe
Lila Thornton
Jason Unrine
Mark Wiesner

CEINT
Duke University
VirginiaTech
Carnegie Mellon
Baylor University
UK University of Kentucky
Howard University
Stanford University

us-eu bridging nanoEHS research efforts

Lisa Friedersdorf
Stacey Standridge
Treye Thomas
Georgios Katalagarianakis
Tom van Teunenbroek
Sally Tinkle

Geert Cornelis
Iseult Lynch
Tassos Papadiamantis
Marianne Matzke
Eva Valsami-Jones
Claus Svendsen
Lee Walker
Melanie Auffan
Jean-Yves Bottero
Camille de Garidel-Thoron
Armand Masion
Jerome Rose
Nina Jeliaskova
Egon Willighagen

eNM eNanoMapper
NanoFASE
Serenade
Mod Nano Tox
rivm National Institute for Public Health and the Environment
NanoMILE
NanoCommons Nano-Knowledge Community
ProSafe
NANOREG

Stacey Harper
Nathan Baker
Mervi Heiskenan
Iseult Lynch
Luisa Russell
Sharon Ku
Mark Tuominen
Mark Hoover
..and so many more

NCIP National Cancer Informatics Program