NNCO 2019 NanoEHS Webinar Series
November 12, 2019

Global Harmonization of Nanoinformatics:

A Case Study in Convergence and Team Science

Dr. Christine Ogilvie Hendren
Executive Director, Center for the Environmental Implications of NanoTechnology (CEINT), Duke University

Dr. Fred Klaessig
Manager, Pennsylvania Bio Nano Systems

Welcome by Dr. Stacey Standridge
Deputy Director, National Nanotechnology Coordination Office

>> Stacey Standridge: Good afternoon. Welcome to today's webinar. I'm Stacy Standridge, Deputy Director of the National Nanotechnology Coordination Office (NNCO). Thank you for joining us for this final webinar in the 2019 NanoEHS Webinar series.

The series examined areas of significant progress in nanoEHS research and was among the activities marking the 15th anniversary of the authorization of the NNI. For another perspective on nanoEHS scientific advances in the last 15 years, you may also be interested in a Nature Nanotechnology Commentary published last week, authored in part by Lisa Friedersdorf and Rhema Bjorkland from our office. I actually have to step off about halfway through the webinar today, so Rhema is also going to skillfully handle the Q&A portion of the webinar at the end of this discussion.

Today's presentation will offer insights into how teams of informaticists, EHS scientists, and computational experts have advanced the potential for informatics to support EHS assessments and safer-by-design capabilities.

I am sincerely pleased today to introduce our speakers, Dr. Christine Ogilvie Hendren of Duke University and Dr. Fred Klaessig of Pennsylvania Bio Nano Systems. Fred and Christine have been among the leaders in these collaborative efforts. Before I turn it over to them, please remember to visit nano.gov for more information on upcoming webinars. You can also follow us on Twitter (@NNInanonews). Fred and Christine, thank you very much today for your time. I'll turn it over to you.

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

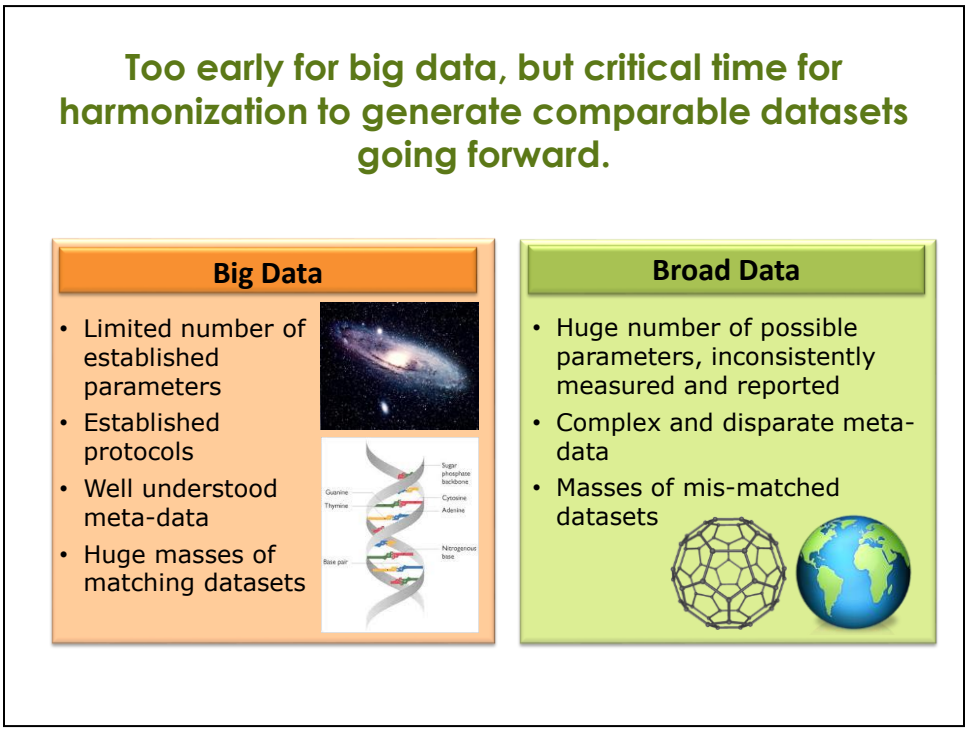
National Cancer Informatics Program

us eu
bridging nanoEHS research efforts

>> Christine Ogilvie Hendren: Thank you so much. I am sorry to say that after that very smooth introduction, I have messed up the display; my apologies!

In the meantime, I am really pleased to be the one to kick off the final in this webinar series. Fred and I have worked together over the past 8, 10 years. An interesting tidbit, which you will see kind of threaded through this talk once I am able to get started, is that we have never been paid to work together, so that's kind of a perfect control for the type of organic convergence that needs to happen across these types of fields; it's because of aligned incentives and scientific goals, and shared research hopes that we can work together.

We decided to title our talk "Global Harmonization of Nanoinformatics, A Case Study in Convergence and Team Science." As we've looked back over these past, you know, 8, 10, 15 years, depending on how we've all kind of gone through this, we really realize that not only have we learned a good deal about nanomaterials and their behavior in applications and in the environment, but also about how to do this type of convergent work where we're asking fundamental science questions at a multitude of different scales of complexity, of size, and of impact, at the same time as we are wanting to support decisions about these. So, kind of doing that in tandem has been a real pleasure and one of the big series of learnings.

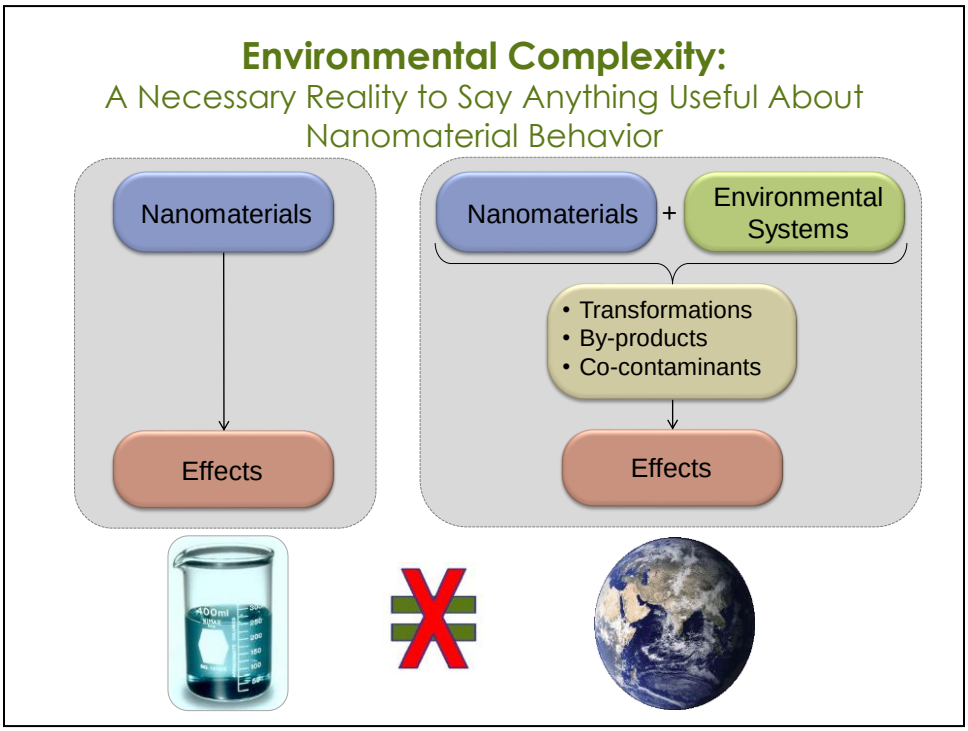


>> Christine Ogilvie Hendren: The first point I want to make is about nanoinformatics in general. We throw around the term nanoinformatics in a time when we're talking about big data a lot; everybody wants to kind of harness the potential for big data to answer big questions—these multiscale questions that we're talking about.

It really is the case that for nanomaterials, we're not at a point of having big data so much as broad data. So it's a critical time to harmonize, even if we're not going to be asking these large quantitative questions.

For example, big data would play a role in fields like astronomy and genetics, where there's well-understood metadata and huge masses of matching datasets. As opposed to what we have in nanomaterials, by and large, which is a huge number of possible parameters inconsistently measured and reported. There are complex and disparate metadata that are required along with these data, and masses of mismatched datasets.

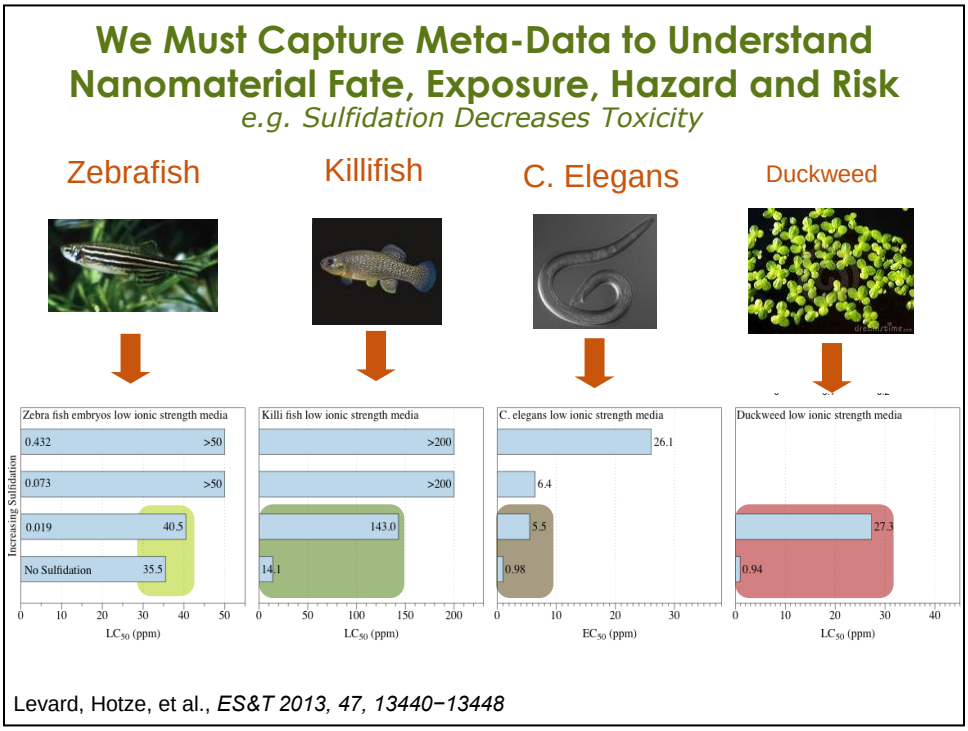
So we have years and years of knowledge, and it's really important. I know many people on this call cannot tolerate, much like myself, hearing people stand up and say we don't know much about nanomaterials. The time for that was decades ago. We know a whole lot. But there is work to be done to figure out how we can match together a lot of these different datasets.



>> Christine Ogilvie Hendren: Something that we have to do in order to do that—and I really appreciate you all bearing with me, pretending we're more at a philosophy talk than an engineering talk, which we are so used to relying on Powerpoints for—But as we think about how to tackle this broad data, that are disparate datasets, one thing that we have to kind of roll into our understanding is that there's just a ton of environmental complexity that is a necessary reality in order to say anything useful about nanomaterial behavior.

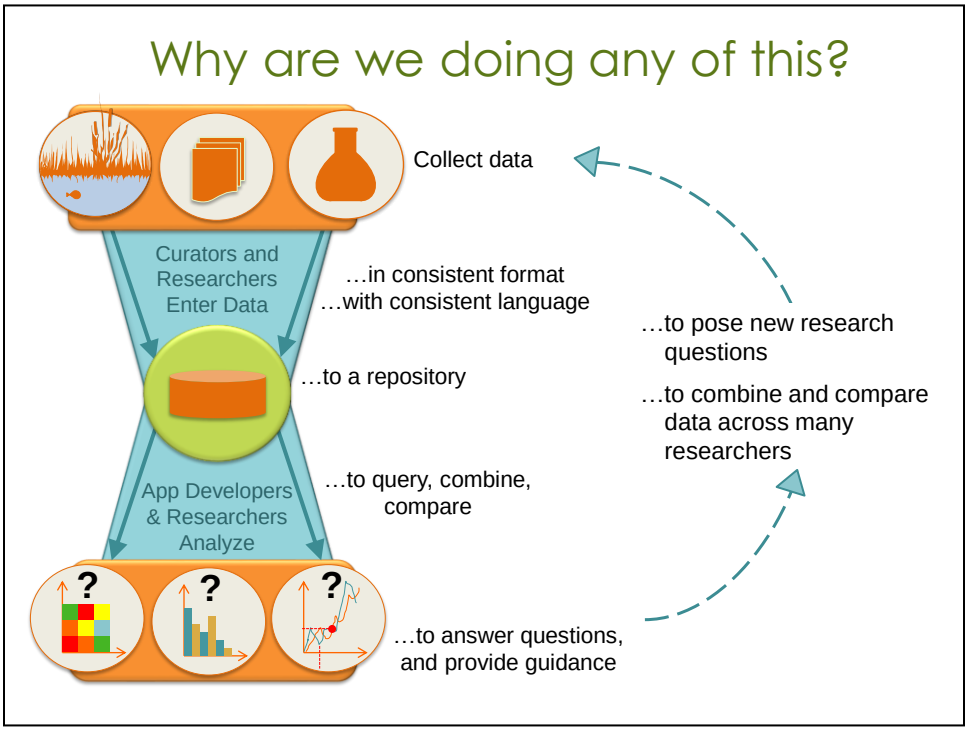
Where we had our arms around a little bit of a chemical sort of fate and transport modeling, or chemical risk assessment, that differs from nanomaterial risk assessment. You can't just go, as you might want to, from nanomaterial effects alone.

Here, I'll pause to catch up with myself. So this is my point—better harmonize, but we don't have big data. And we need to go not just from nanomaterials to their effects, as you might in the lab, but you have to consider nanomaterials as part of their environmental systems, including all the different things that may happen to them as they interact with each other, with the environment around them, and then the effects that you might care about. In the real world, we have to roll in some of this complexity, if you are going to make useful statements.



>> Christine Ogilvie Hendren: This is an example of the many papers that came out illustrating this, which is that without the metadata about the environment of a nanomaterial, its initial state characterization is not going to predict very well the toxicity.

This is a paper from Greg Lowry's group at Carnegie Mellon that came out of CEINT (Center for the Environmental Implications of NanoTechnology). It shows that the same nanosilver in different environments is going to have much different toxicity. Sulfidation decreases toxicity. If you try to predict the outcome based on the initial state, you would be wrong. You've got to consider a lot more. You've got to consider a lot more interactions and dynamic systems.

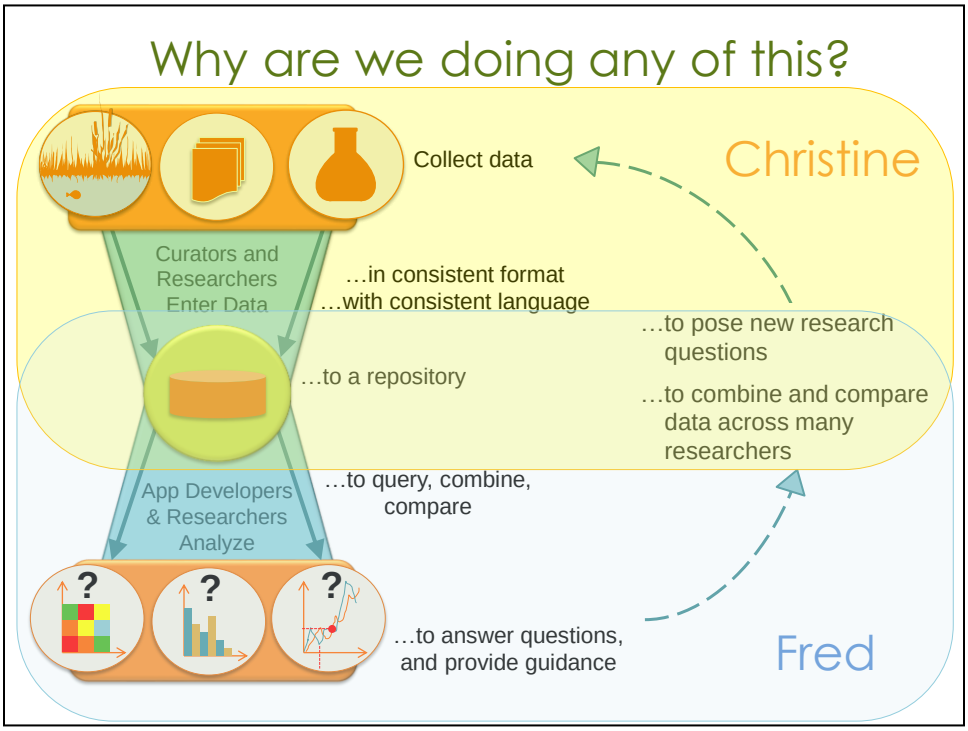


>> Christine Ogilvie Hendren: So, an important thing to realize as we look for informatics approaches is why we are doing any of this.

This seems obvious, but it's quite easy to lose track of that as a compass when we're in the middle of organizing data. We're doing it to answer questions and provide guidance in real time.

In order to do that, we need to query and combine and compare datasets. We need to get them from a repository. We need to have that enabled by consistent formatting and language. And we need to collect data that would feed into all of these. So the reason we do that backwards is we have to back ourselves into a scenario where we can compare our disparate datasets.

If we do that in the right order, then we can get better questions together and provide guidance and pose new research questions that are even more harmonized in terms of being able to compare the measurements.



>> Christine Ogilvie Hendren: Today, I'm going to talk about, loosely, this top part, really collecting the data, curating it—and some history around that—into consistent formats. And then Fred is going to take over and talk more about what do we do with it, what's the fun stuff we can ask.

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.

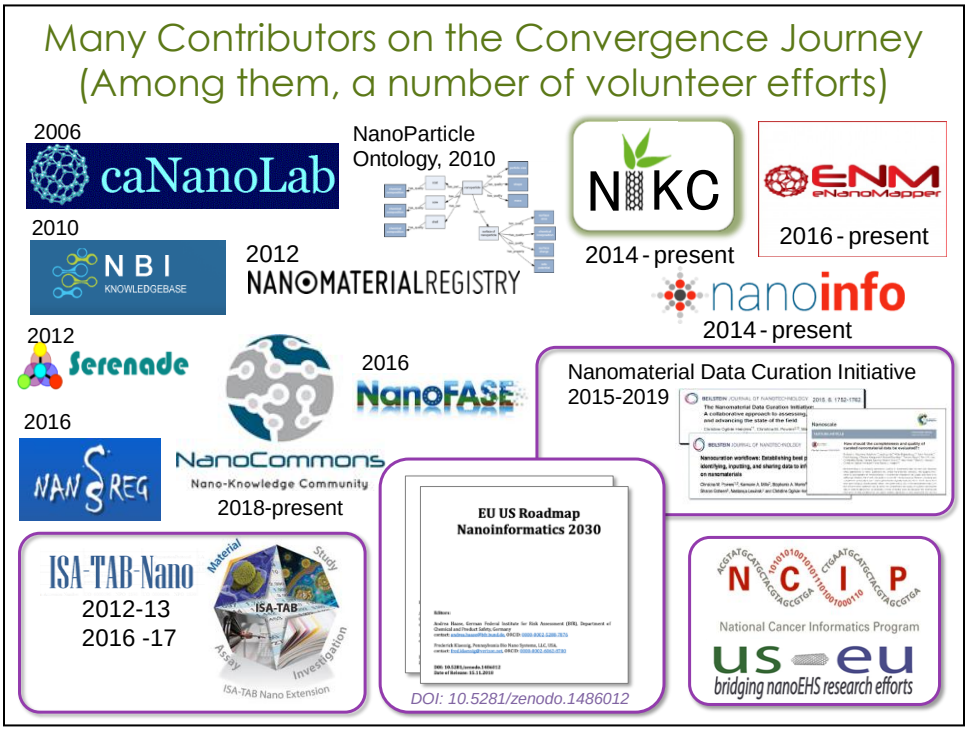
If we do it right, we can even use it for things down the road that we can only imagine now.

>> Christine Ogilvie Hendren: This is just to make the point that avoiding divergence in all of this takes effort, consistent contact, and commitment and patience, often outside the paid project, just because of the way things are funded.

So divergence in cyberinfrastructure development makes sense because each project is separately funded, separately incentivized, and has a finite pot of resources.

We're seeing all these different pots within a greenhouse. There’s only so big this tree can get. So we want the resources to grow together, make each other stronger, leverage each others’ “nutrients”, and be something wonderful together. Doing much more collectively than we could ever do individually. If we do this correctly, we can be like this guy, and even use it for things.

So that's what we want to do: Have the foresight, while we're individually working on our separate projects, to converge together, and that takes some investment.

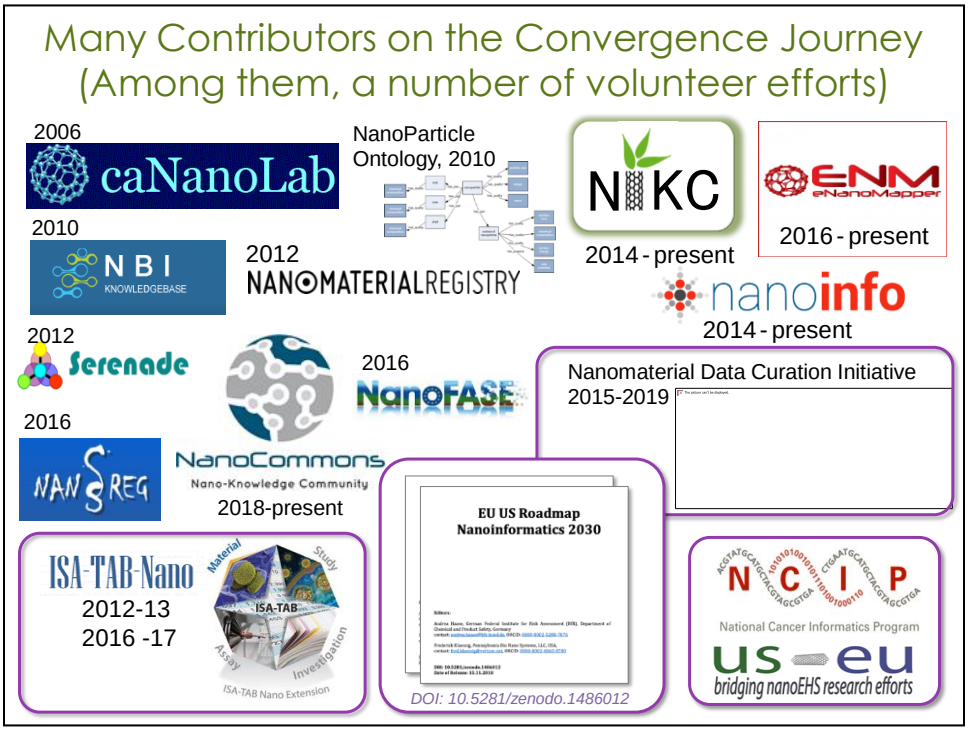


>> Christine Ogilvie Hendren: This is a very—I need to note—*incomplete* picture of some really key pieces of the puzzle in contributing to the convergence journey. And it's notable that among them are a number of volunteer efforts.

So caNanoLab started out at NIH (National Institutes of Health) in 2006, really saying, okay we've got to pull together our data across many different projects in order to compare.

NanoParticle Ontology (NPO) came out of PNNL (Pacific Northwest National Laboratory) and through the Nanotechnology Working Group, which is part of the National Cancer Institute (NCI) National Cancer Informatics Program (NCIP. That sort of begat not only the NanoParticle Ontology, but took advantage of the NBI (Nanomaterial-Biological Interactions) Knowledgebase, which one of the NCI team's Nanotechnology Working Group's long-time chairs, Stacy Harper, donated—it's 148 perfectly matched Zebrafish results. So starting to get some critical mass.

Several projects, including the Nanomaterial Registry, started to look at what are these minimum characteristics. And then you see the Nanoiformatics Knowledge Commons (NIKC) from CEINT and nanoinfo, which is the database and toolset from the sister center, UC Center for Environmental Implications of Nanotechnology (CEIN). Both of those created some infrastructure. (*cont.*)



>> Christine Ogilvie Hendren: (*cont.*) The eNanoMapper project, in Europe, Serenade, NanoFASE, NanoCommons, NANoREG. You're seeing tons of different European projects starting to put together both their datasets and also create some infrastructure.

The next important point to make is really that these four on the bottom, ISA-TAB-Nano (Investigation/Study/Assay, tab-delimited format for nanotechnology data); EU-US Nanoinformatics Roadmap—not only the 2030 but the 2020; the NCI Nanomaterial Data Curation Initiative, which is a series of papers over four years; and then of course, the NCIP (Nano Working Group) and the U.S.-EU Communities of Research. These are all supported by excellent administrative support and organization, but project-wise, they were not funded. And so it really speaks volumes about the shared incentives and the understanding of the community of what was needed to coalesce and really groom these “potted plants” together, that it behooved everyone to volunteer their time in order to make more together than we could individually.

>> Christine Ogilvie Hendren: I'm going to go now into just a very brief abridged history of building the NanoInformatics Knowledge Commons. This is because it's the project with which I am the most familiar, but it is not to say that this is the primary work in the field. It's just illustrative of what's required to grow together and get from this disparate data to one day being able to be big data.

This is built off of the concept of the old English Commons. Everyone can put their sheep out to pasture to take advantage of the open fields and feed their sheep, and if we are all tending the fields and making sure that we are taking care of them, then it should work out. What we don't want is a tragedy at the commons, right? So we'll say that the NanoInformatics Knowledge Commons is supposed to be for the greater good, a place where nanoinformatics datasets can be placed. Some of the earlier projects from that brief history slide before that did nothing but "lay down grass" really put a lot of work into saying here are some tools, ontologies, language, dataset formats, or Nanomaterial Registry, eNanoMapper, and NPO and ISA-TAB Nano projects of the Nano Working Group. This laid the groundwork for being able to start to put some grass down in these fields, so to speak.

(*cont.*)

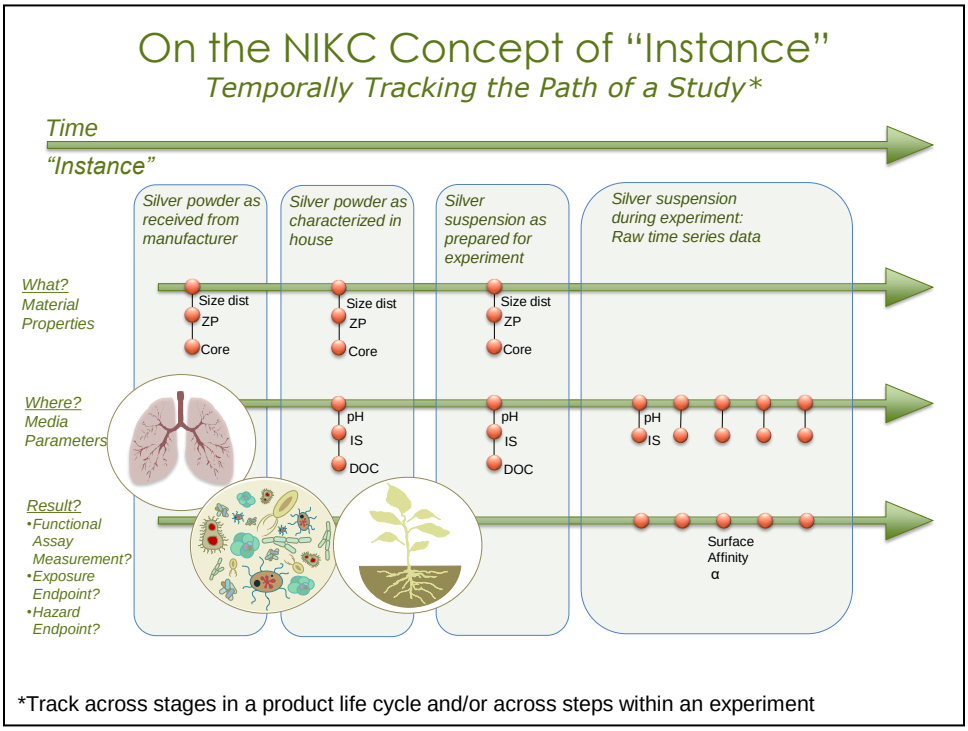
.

>> Christine Ogilvie Hendren: (*cont.*) CEINT is one of the projects that then tried to say okay, how can we not only lay down some more infrastructure, language, and terms, but also then start to—funded by particular efforts, including the Consumer Product Safety Commission and Army Corps of Engineers—start to then put some sheep on this so that they can begin to graze?

NanoFASE did the same thing; Serenade did the same thing. These are European projects that were taking advantage of the infrastructure laid by others but then also saying hey, we're going to devote maybe a whole postdoc's worth of time to taking care of this.

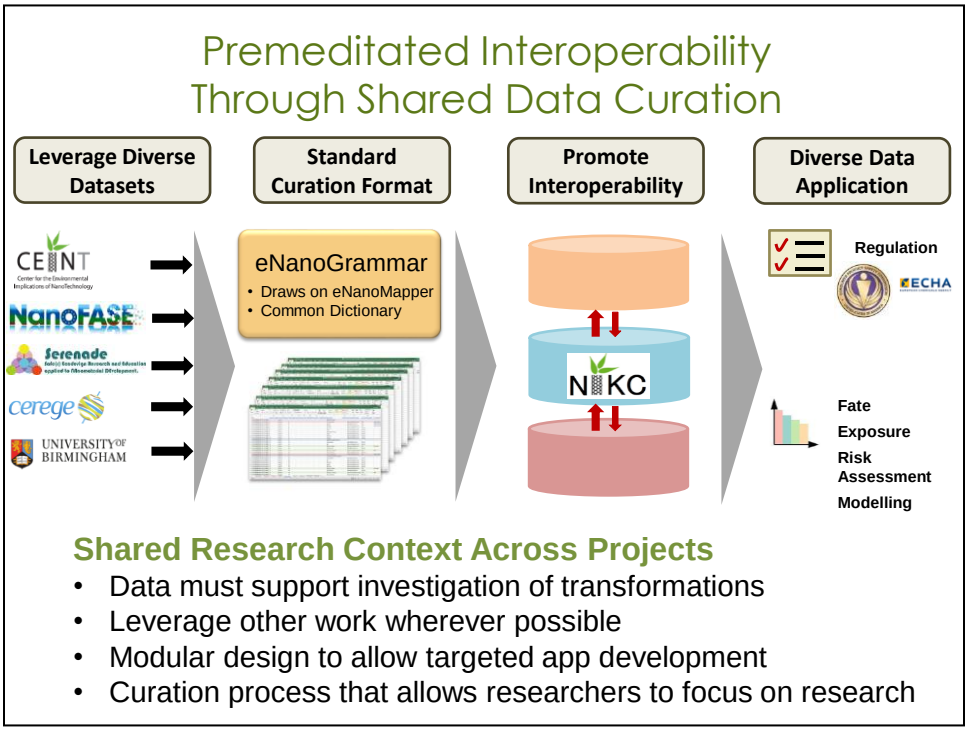
NanoCommons is moving along and doing the same thing, and advancing that, adding in eNanoMapper, and other types of ontologies that are really kind of snowballing, so that we get critical mass. Now, it is basically time for other people who deem it useful to collect their data at this granularity and in this format to add their datasets and then also “graze”, where grazing is the analogy to querying and making use of these data.

.



>> Christine Ogilvie Hendren: The NIKC concept of "Instance". What is special about this in particular, this database? There are many other ways to do it. One of the things that brought us all together, the folks that are putting in our datasets to this format (which is a bit laborious) is this realization of the slide that I showed earlier with the sulfidation, showing that changes in the characterization of a material can depend on the media and the surrounding characteristics. If you want to describe a nanomaterial's journey and its transformations, which are, as we covered, important in order to tie anything to effects you might predict, you've got to not only characterize it as maybe you receive it from the manufacturer, but also tie those characteristics to the media and the surrounding parameters as it changes over the course of an experiment…

Or, as Fred will talk about more, over the course of the value chain, as it changes hands across its real journey through its life cycle. So these surrounding media parameters could describe the inside of a human lung, could describe a microbiome, or could describe a rhizosphere. All of these could be compared if we are capturing the data in ways that are conducive to them looking at comparisons between datasets that maybe we didn't even intend when we took the dataset.



>> Christine Ogilvie Hendren: So this is what we call in the Nanomaterial Data Curation Initiative paper "premeditated interoperability." That's through shared data curation, and this is just showing how you can have a shared research context across projects where you leverage diverse datasets. Here I just show the projects that are in this particular collaboration: CEINT, NanoFASE and Serenade in the EU, CEREGE, and the University of Birmingham (see NanoCommons data and metadata curation).

These are all devoting time, and that's the important investment of postdocs, in order to enter their data into a standard curation format using eNanoGrammar—that means you have the same format, same ontology, same terms—and then putting them into interoperable repositories, even if you have more than one. If they can interchange their data, that's good. That allows you to apply this in diverse ways, which Fred will cover much more, but to support regulation and more science.

So in this way, you leverage work, wherever possible, that's already been done. The NanoCommons and NanoInformatics Knowledge Commons did not seek to make up new terms. Anywhere we could use eNanoMapper terms, we did. Anywhere they could use nanoparticle ontology terms, they did—so that we're always going in the same direction. And the curation process, if you design it modularly and invest in curators who help to shepherd this information, allows researchers to focus on research.

>> Christine Ogilvie Hendren: This, I'm going to go over very briefly just because we got started a little late due to my technical snafus. This is something I heard out of Northwestern Team Science and Nora Savage at NSF, which is, convergence can be described as sort of the smoothie of this diagram. I think it's important in thinking about nanoEHS, because convergence isn't *better* than disciplinary work, so to speak, but it is *fit-for-purpose* for things that are as complex as nanomaterial environment, health, and safety research.

So you might need disciplinary research: If you're going to rethink colloid theory, you need colloid scientists. Multidisciplinary research might be people in the lab handing off to social science colleagues to do something separately, as a baton pass.

Once you get to transdisciplinary research or convergence research, this is inextricable mixture. This is an ecologist asking a new question because they've been exposed to synchrotron collider techniques, so they can ask new questions—say, about speciation in the sight of an earthworm—that before they worked together closely they might not have been able to ask. Diverse teams are necessary for this, and we saw this time and again in nanoEHS research.

Three Keys to Enabling Convergence Exemplified by Nanoinformatics for NanoEHS

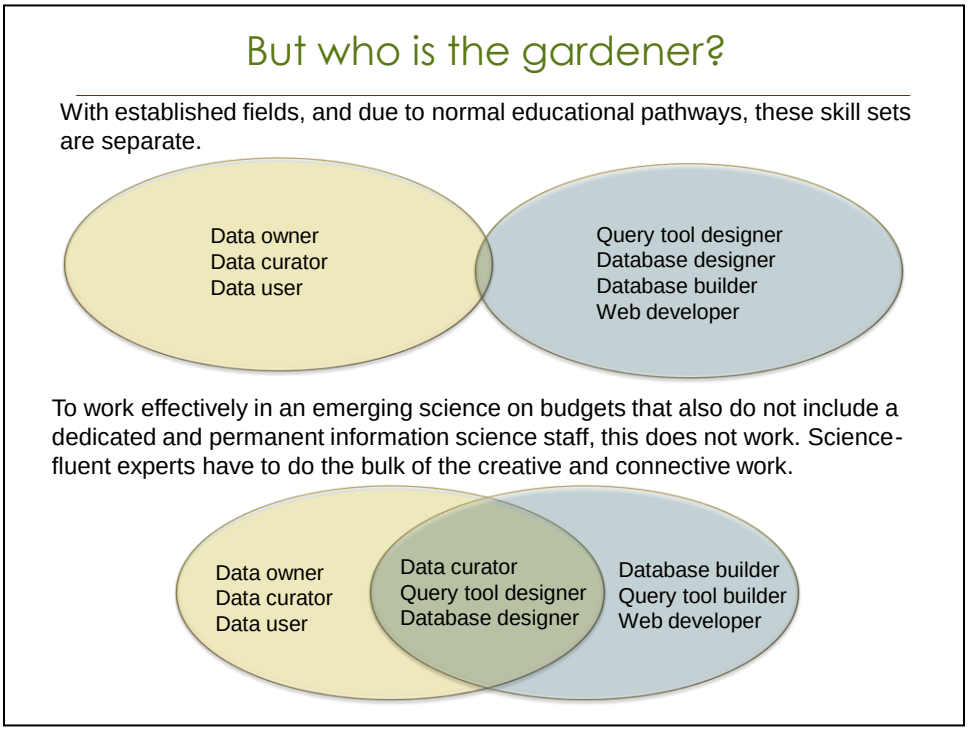
1. Integrate research approaches and methods through **a shared purpose** that promotes harmonization of methods[1], media[2], measurements, language
2. Shepherd research-driven **data** integration
3. Invest in Integration and Implementation Science (I2S) **specialists** to coordinate **teams**

1 – A functional assay-based strategy for nanomaterial risk forecasting, https://doi.org/10.1016/j.scitotenv.2015.06.100
2 – Harmonizing across environmental nanomaterial testing media for increased comparability of nanomaterial datasets https://doi.org/10.1039/C9EN00448C

>> Christine Ogilvie Hendren: Three keys exemplified by nanoinformatics for nanoEHS that enable convergence are having a shared purpose that promotes *harmonization of methods*, for example—and here we call out the functional assay method that we were able to introduce partway through the nanoEHS journey; *harmonization of media*– so if we're going to test a nanosilver particle, for example, let's test it in similar media so then we can compare, and this points to another paper (cited in *footnote 2* on slide*)* where many of us tried to work to decide some of those; and *measurements* and *language. So* we're all kind of swimming, again, towards *harmonized datasets*.

We have to shepherd, again, research-driven data integration. Going back to why are we doing this is to answer questions, not just to organize our datasets. That is a nontrivial finding that helps us filter what we do and what we don't.

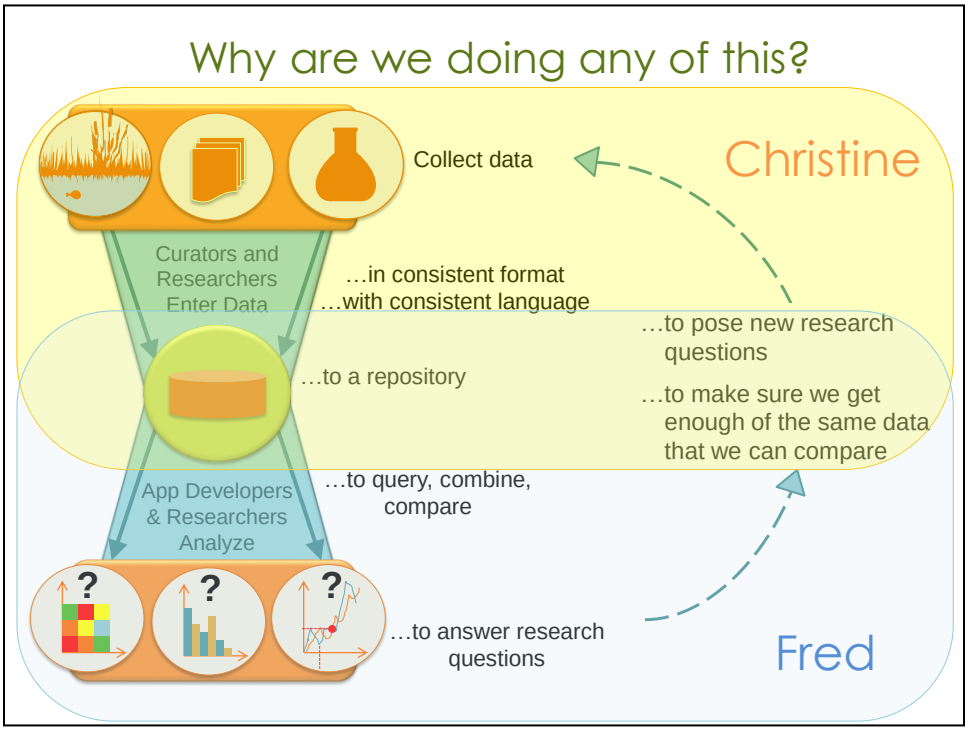
And then, as alluded to a couple of times, investing in integration and implementation science specialists, people who coordinate teams. Importantly, I've realized that the blender making this smoothie needs a lid on it. We can't just combine everything together and ask all the questions; we have to use these specialists, the data, and the shared purposes to point us to and put a limit on all the different things we open up and ask.



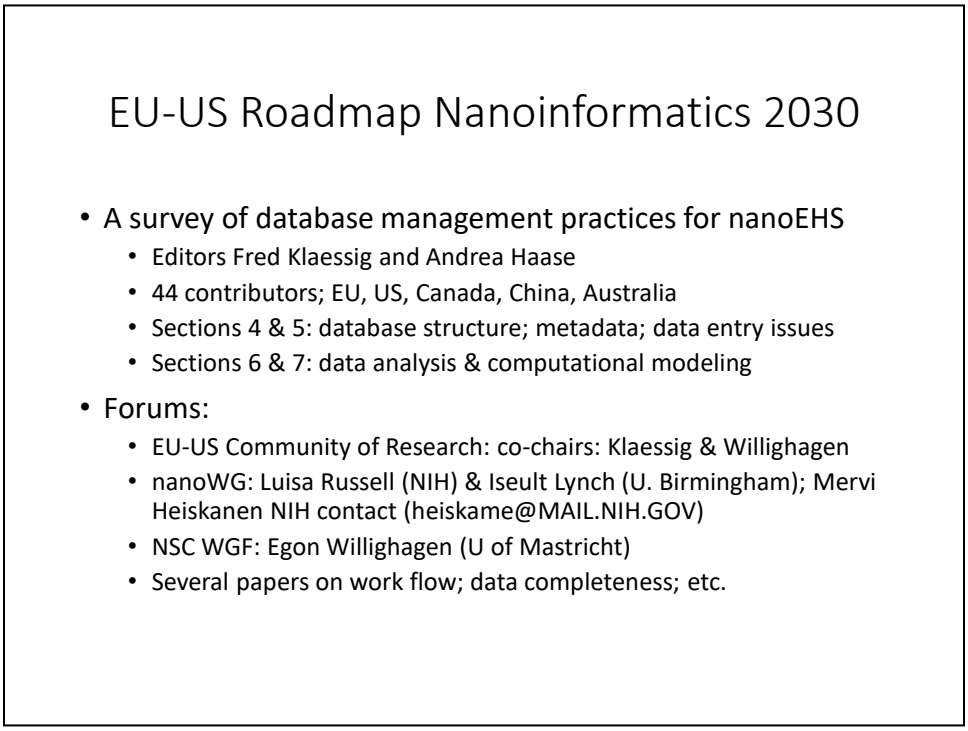
>> Christine Ogilvie Hendren: So who is the gardener?

With established fields, maybe more the astronomy and genetics that I've talked about, you may have data owners, curators, and users in one camp and then the people doing what we would call the IT side in another.

But to work effectively in an emerging science, on things that do not include giant, Microsoft-style budgets, we can't do it this way. We have to have science-fluent experts in the middle who have to do the bulk of the creative and connective work between these sort of spheres, while we're having disparate data and hoping to one day lay the tracks to have interoperable big data.



>> Christine Ogilvie Hendren: At this point, I'm going to hand it off to Fred to talk about why are we doing all of this and what fun things can we get out of it?



>> Fred Klaessig: Thank you very much, Christine, and hello, everybody.

Christine and I share many overlaps in our presentations in terms of community involvement, Communities of Research, and mutual dependence within the community on making contributions. My perspective from an industrial background makes me wonder where this is taking us, especially in terms of future regulatory requirements or addressing future regulatory considerations. And the other area that's important to me is what the value is and where will standardization promote this effort. That comes from my being on the E56 Committee of ASTM, where I'm chair of the Informatics and Terminology Subcommittee.

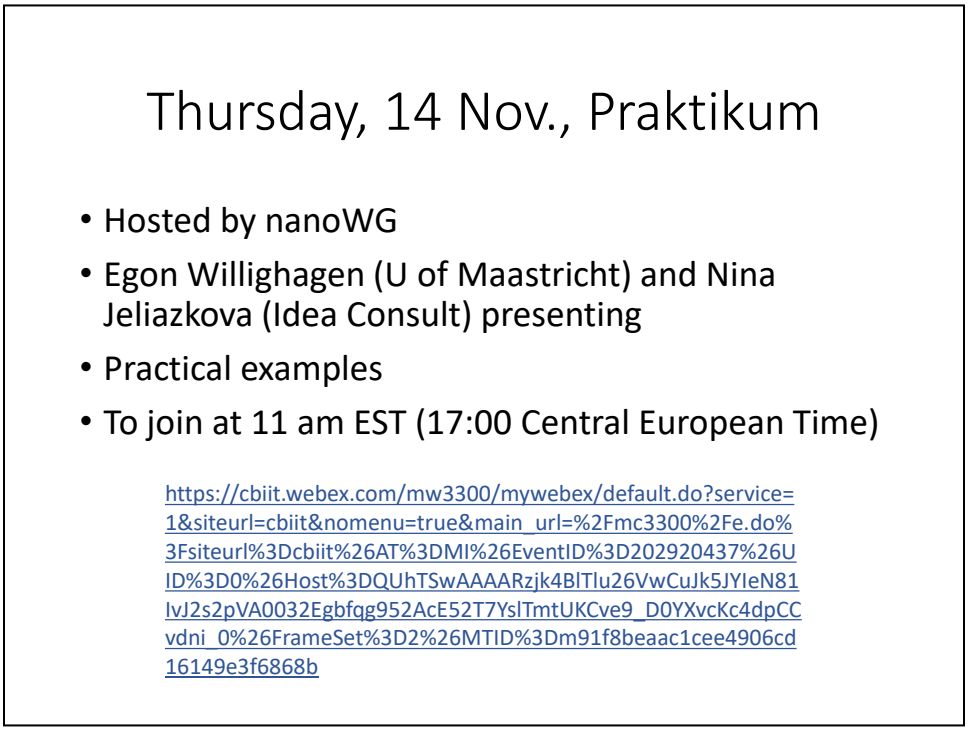
For me, the starting point is very much the EU-US Roadmap Nanoinformatics 2030, which I co-edited with Andrea Haase of the German Institute for Risk Assessment. It's rather long, but it does have the issues. It had 44 contributors, many from the EU, U.S., some from Canada, China, and Australia. What Christine has been mentioning has been primarily sections 4 & 5, database structure, metadata, getting the data in, the need for the community to have agreements to do that. My presentation will be more on sections 6 and 7, on getting the data out: What is analysis? What does computational modeling mean? (*cont.*)

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 Russell (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.

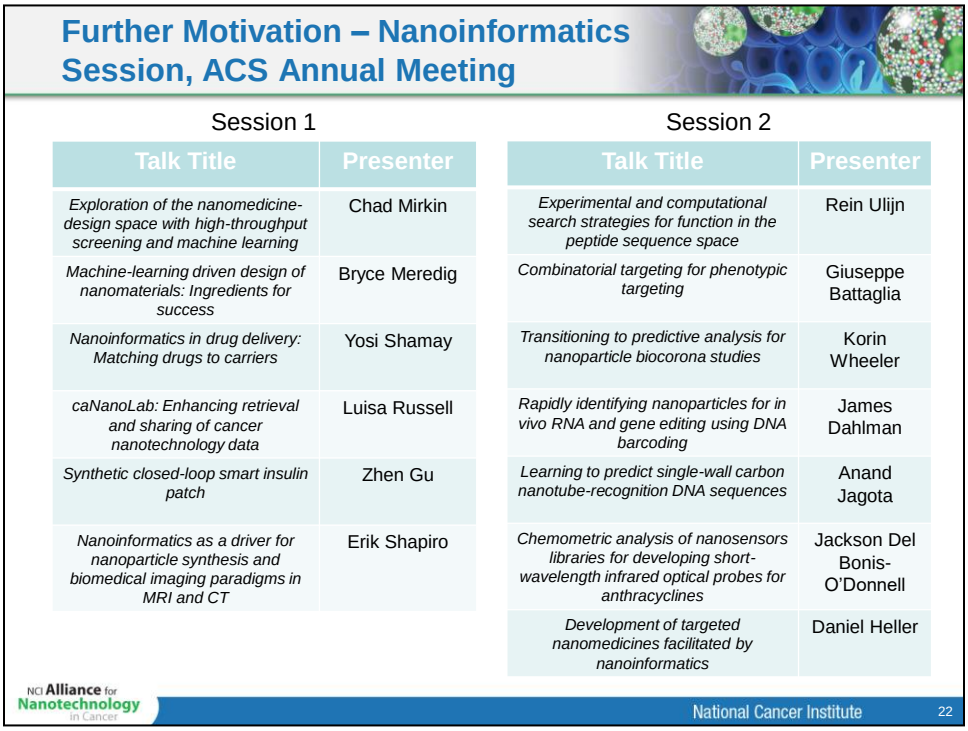
>> Fred Klaessig: (*cont.*) We've already discussed that there are a number of forums that are ongoing that pursue this. There's the EU-U.S. Community of Research that I co-chair with Egon Willighagen of the University of Maastricht. There's the Nano Working Group with Luisa Russell of NIH and Iseult Lynch of the University of Birmingham, and they have a weekly or almost-weekly teleconference on Thursdays. Mervi Heiskanen is the contact if you wish to get on that mailing list. And the NanoSafety Cluster in Europe has the Working Group F (Data Management), led by Egon Willighagen.

These groups have issued several papers; I've put some references in one of the slides at the end of the presentation. In general, the U.S. colleagues are working on workflow, data completeness, and issues of community. The colleagues in Europe have tended to focus on the functionality issues that are important to generating the database.

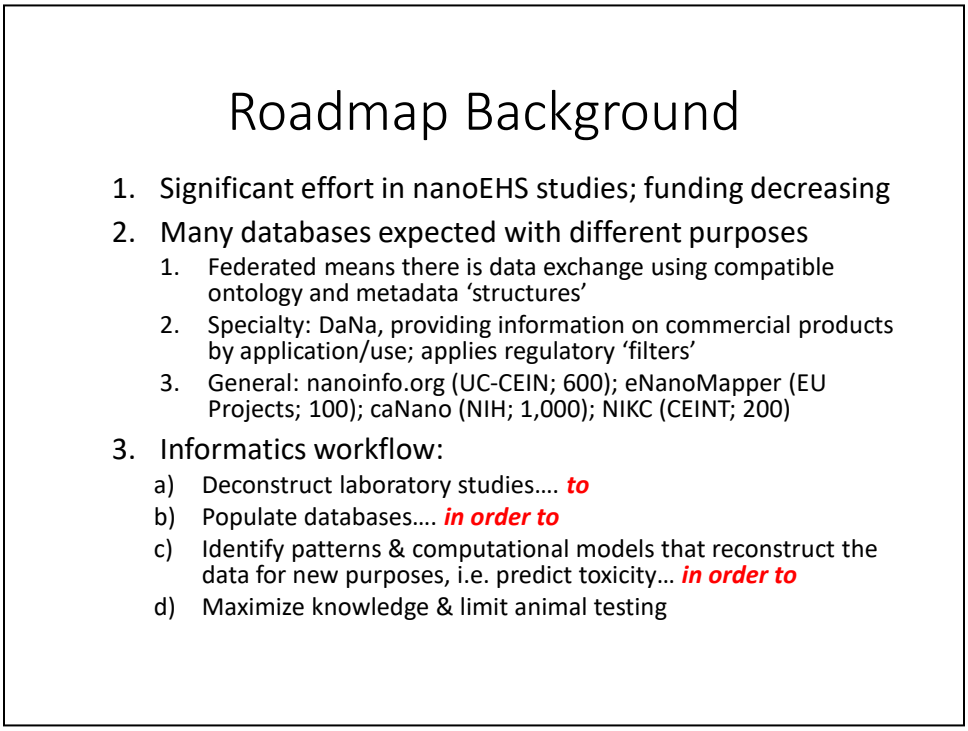


>> Fred Klaessig: Just as an example of the coordination, as Christine said, we're doing this on a volunteer basis by our interests. But the Nano Working Group will have a presentation this Thursday where Egon Willighagen and Nina Jeliazkova from the European side will give some practical examples of what Christine and I are discussing from a more philosophical side.

I've put the long form of the connection here to the meeting. I think all of you have also been sent an active link by Rhema in your invitations.



>> Fred Klaessig: Additionally, just again to see that this is a growing topic area, this is from Luisa Russell; it's the agenda for a recent ACS (American Chemical Society) meeting that had a nanoinformatics session. You'll see Luisa's name there on the left side, and Chad Mirkin's. This is really an activity that is more on the nanomedicine side, whereas Christine and I generally discuss it more on the nanoEHS side in terms of environmental toxicology.



>> Fred Klaessig: Going back to the roadmap, there's been a significant effort in nanoEHS study. The funding is decreasing overall, though perhaps in certain areas there's a focal point. There are many situations here in the United States and in Europe where people are trying to merge existing databases, but we also expect that local databases will continue. Where there's a local need, local maintenance, there will always be a need for some level of federation, harmonization, some ability to have a compatible ontology, and metadata structures among them. Some of these databases may be specialty.

I've highlighted here DaNa, which is a database maintained by the German chemical industry along with their colleagues in government. It provides information on commercial products by application and use. It also applies what they believe are appropriate regulatory filters, so that one is getting what would be acceptable regulator-type information on these products.

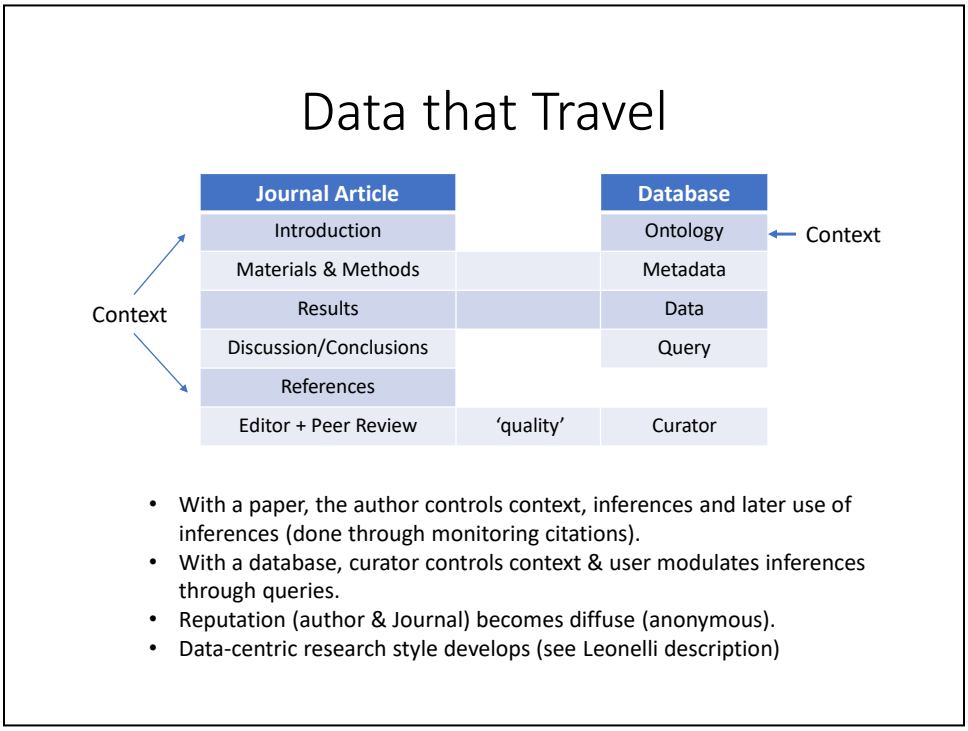
There will additionally be what I'm going to call more general databases. These would be those that involve research-laboratory types of organizations: nanoinfo.org is from the UCLA-headquartered Center for Environmental Implications of Nanotechnology (CEIN); Christine has discussed the NIKC; Egon Willighagen is instrumental to the eNanoMapper; and colleagues at NIH (Luisa Russell, Mervi, and others) are involved with the caNanoLab, which is nanomedicine. (*cont.*)

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

>> Fred Klaessig: (*cont.*) I've put here as a poor metric the number of discrete entries in these databases to my knowledge; that's a poor metric in that each of these databases has a different purpose. The one at UCLA pursues modes of toxicity and therefore has a number of materials in it that help them demonstrate certain adverse effects in their mechanisms. eNanoMapper is closely associated with the European Chemicals Agency, and the caNano group is more the therapeutic applications.

My expectation would be that having some familiarity with at least two of them will be appropriate in the future. All of them share an informatics workflow. They deconstruct laboratory studies in order to populate databases, in order to identify patterns or to pursue hypotheses through the computational models, in order to reconstruct the data for a new purpose, which generally is predictive toxicology. And that in turn maximizes our knowledge and serves to limit unnecessary animal testing.

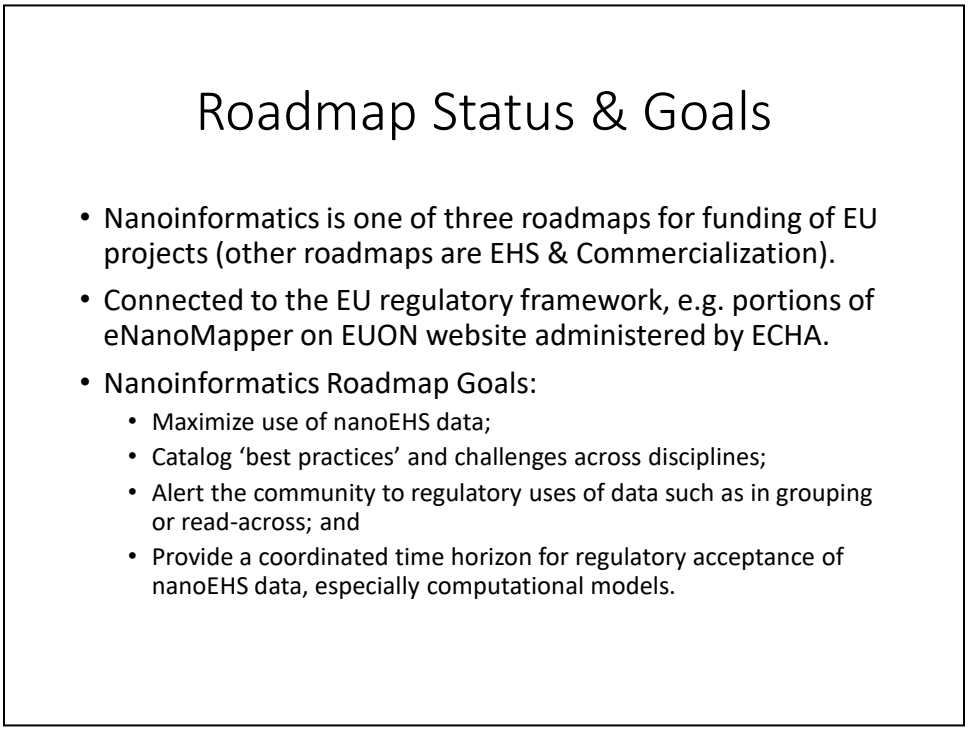


>> Fred Klaessig: That deconstruction can run counter to or has some alignment, shall we say, with a traditional journal article.

I've put on the left the journal article section, on the right the database. Basically, the results of a journal article become the data in a database, the materials and methods section become the metadata. At the bottom of the left side is the quality system, that is, editorial review, peer review, which now moves over to the curator function, which is a new function. That's the person who's going to annotate the results of the metadata and also serves presumably as some sort of decision-maker as to whether something enters the database or not.

What would normally be the context, introduction, and the discussions and conclusions becomes in this case the ontology, very often controlled vocabularies. That's an area where E56 has been involved with the recent standards on the uniform description system from John Rumble and others. There's also ISO TC 229 in terms of providing definitions. But that provides the context. What would normally be inferences drawn by authors become the query or becomes the user trying to see if they can repurpose this information for their direct intents.

I propose that there's going to be a different type of research style as one develops this, and I'd like to give some examples of where that is going.



>> Fred Klaessig: Going back to the nanoinformatics roadmap to start, it's one of three roadmaps that are used for funding EU projects—there's one in EHS (lead author was Vicki Stone), and there's one in commercialization. You can find one of the results, the eNanoMapper, on the EU's Observatory for Nanomaterials website administered by the European Chemicals Agency.

Again, the goals are maximize the use of nanoEHS data; catalog best practices, but also catalog challenges that cross disciplines, picking up the community issues involved; alert that community to the regulatory uses of such data as in terms of grouping or in terms of read-across; and lastly, to provide a coordinated time horizon for regulatory acceptance. In this case and my interest in particular, it's only after the regulators have come to accept some of the new developments that we will be able to take advantage of all the work that we've actually been doing.

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)

>> Fred Klaessig: I'm going to emphasize computational models. There are models that are already used when one goes to the EPA, which is my experience. They determine an analog to the new submission and the analog would have a more complete dossier of information that allows them to use their best knowledge on the new submission.

The analogs, very often determined by QSARs, which are quantitative structure-activity relationships, are computational. OECD (Organisation for Economic Co-operation and Development) has criteria for these models—a defined end point for one is a biological test. There should be clarity so that people understand what the algorithm does and whether it's statistically robust. The criterion I like the most is the mechanistic interpretation, if possible, because there's a recognition that in this area, one can often find correlations but not always causation; therefore, where possible, a mechanistic interpretation is desired.

Because very often the evaluations by regulators are confidential, it becomes difficult to get regulatory feedback on models; plus, it's their position that they wish for scientists to determine what's important in mechanism and not for the regulator to create that.

I'm going to use an example of dissolution of a sparingly soluble particle. I chose silver. The ion is known to have toxicity, the particle has toxicity, and the question is, regarding computational models, would QSARs be used? And this new acronym, PBPK: physiologically based pharmacokinetics—it should be toxicokinetics (PBTK).

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

>> Fred Klaessig: This is a way that you can translate the dose to the animal or to the administered dose and actually get it down to an organ dose. And this hypothesis with the computational model acting on that hypothesis. The "**PB**" means you made the human or the animal body into compartments that align with organs: you're modeling the physiology. **TK** is the toxicokinetics, it's the action of physiology acting on the toxicant. **A** is absorption, **D** is distribution, **M** for Metabolism, **E** for excretion.

If done well, the overall exposure can be expressed as a localized organ dose, you can then scale that exposure across types of exposure, species, and lengths of time, and it becomes a mix of kinetics and equilibrium concepts such as the oil-water partition coefficient.

If you look at the overall area of informatics in this topic area, one sees a progression of where there are computational models for determining properties that the QSPR—the Materials Genome Initiative would be an example of trying to do that. There's QSARs for biological activity, PBTK models to get organ dose. And these would fit in with a broader perspective of adverse outcome pathways that are the physiological pathways that are used when an adverse effect actually manifests itself.

One of the difficulties is that particles complicate and challenge what has been the conventional way of doing each of these areas. There's uncertain dose metrics, meaning mass, surface area, number, partition coefficients; oil/water does not apply. Dissolution means you don't have a monotonic decrease in concentration that is often assumed by the models. And adsorption is just a problem for everybody. So I'm going to take an example of a recent NIOSH bulletin on suggesting a new recommended exposure limit for nanosilver.

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

>> Fred Klaessig: They have in their document provided backup information regarding a PBTK model that was authored by colleagues in Europe. What is here in the top left is the Arrhenius coefficient, reaction coefficient, for the ionic material. And on the right is the kinetic reaction rate for the particle material. Each of these is the uptake in an organ. And I want to contrast them as a state of the art, in a sense that this is a significant advance of having closure in terms of the analytical math. But there are also differences. So on the left, they had decided to partition the silver by the glutathione concentration of organs and the body. So you have a mass and you have glutathione, and the fitting constant has the units of inverse minutes.

On the particulate side, the colleagues, for purposes of simplifying and having tractable calculations, decided to use the organ blood volume—that would be the volume flow rate divided by the total blood volume. So you now have a volume, not mass; blood, not body; and the fitting term now is dimensionless. I'm expecting, as a physical chemist, for there to be an energy relationship, an activation energy. I see changes in these constants.

Also, the volume of total blood implies that you are a series configuration—electricity and magnetism, series and parallel resistor groupings—but it is a series and not parallel. So this particulate is allocated by organ blood flow rate, and the ionic silver by fraction of glutathione concentration. The contrast is to show people are working on this, making improvements; at the same time, you see some contrasts that might be improved upon. It also fits in with a coordinated outlook in terms of the physical chemist, and the biologist, and the modeler all needing to be present throughout the process.

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

>> Fred Klaessig: More recently, the European Food Safety Authority (EFSA) has accepted one model. It's called GUTS (General Unified Threshold models of Survival). It's a pesticide model, basically. And they may point to the direction of how regulators will wish to see models presented to them. In this case, there's a framework, definitions, equations, and very important, accepted interpretations. These have to be consistent with implementation, which means when you solve the differential equations or come up with digitization, it has to give back what those accepted interpretations are supposed to be.

The colleagues provided ring data, which is interlaboratory studies. From that, you can select accepted case modules, which basically fit your experimental design to the most acceptable model, or use the modules to design your experiment. And very important is that there's an Excel spread-sheet approach that is web-accessible that allows the regulator or anyone else to examine other scenarios beyond the one provided by the presenter.

Each of these actions is, effectively, shall we say, clarifying what's been called "epistemic opacity," which is another way of saying I wasn't very good at differential equations, I don't know what's in *Mathematica*. At one point in time, one loses sight of what the model is accomplishing in that mechanistic sense. So you see here an attempt to clarify, make certain that one can follow what the modeler is attempting to do.

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

>> Fred Klaessig: In terms of some concluding remarks spanning my own presentation as well as Christine's, a functioning nanoinformatics effort will alter practices on data sharing, analysis, and attribution.

The curator assumes an important role, very often a postdoc in some of the biological sciences, but still to be determined in this area of nanoEHS.

There will be many types and purposes of databases that will take advantage of the ability of a local group to provide local maintenance, but then that puts on a requirement of some degree of federation.

Research consortia should address these database issues at the start. It's easier to collect the metadata if you know what metadata is going to be required.

Regulatory acceptance will be important to industry.

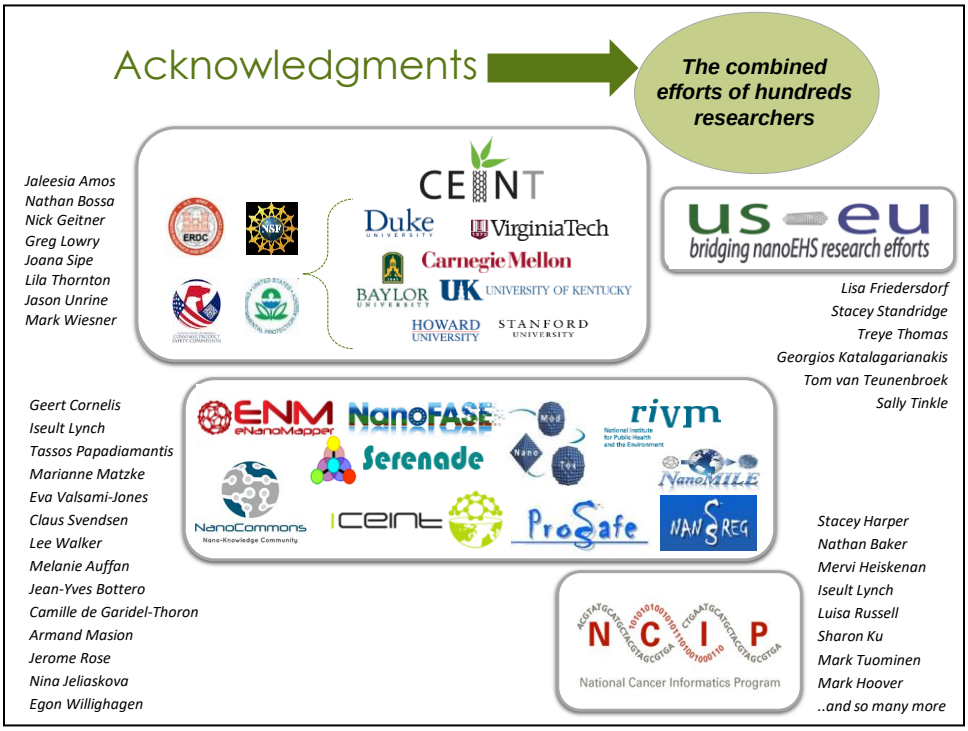
And lastly, for colleagues in the academic and nonindustrial area or nonregulatory area, anticipating how your research data will be recast into a regulatory format would be useful, so once again, you maximize the use of your research efforts.

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

>> Fred Klaessig: I've provided some backup citations, literature, to this that can be, of course, shared.

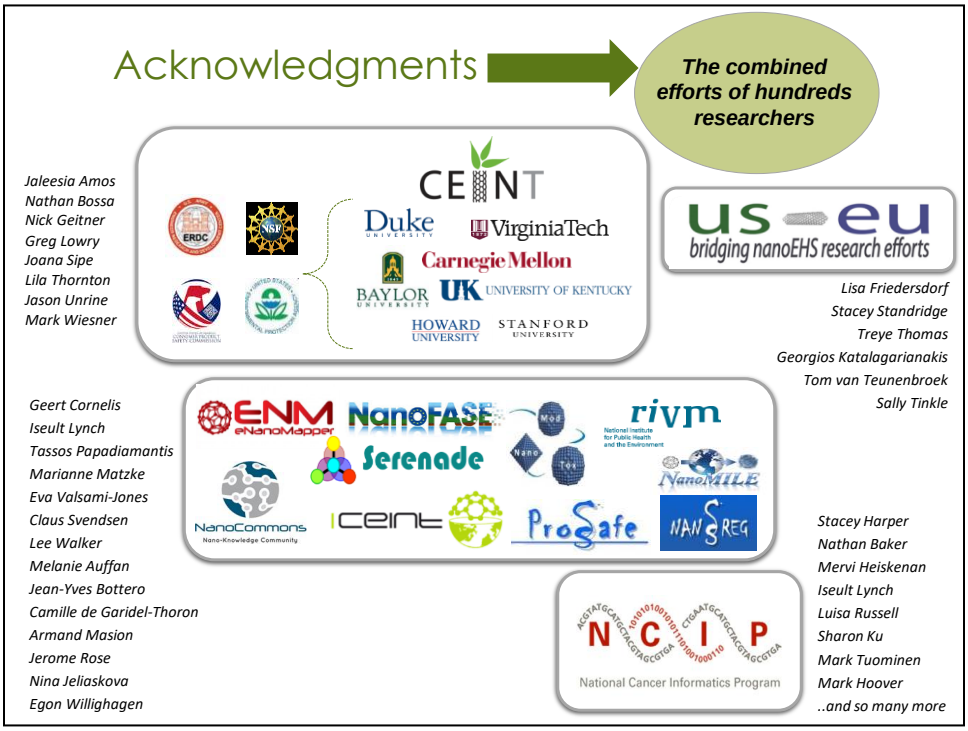
And at this point, I'm going to return it back to Christine.



>> Christine Ogilvie Hendren: Thank you so much, Fred, for a great presentation and handling the "so what" part of this fun journey, saying what have we enabled ourselves to do.

There was no possible way to make an acknowledgment slide that would be sufficient to point out everybody who enabled this, so we're first saying this is a combined effort of hundreds of researchers. Several people are listed here that we highlighted in particular in the U.S. and the EU, and this is by no means, again, exhaustive. I do want to take a second to especially acknowledge the folks who work at the Nano Working Group, which is our shorthand name for the National Cancer Informatics Program Nanotechnology Working Group. This is the volunteer, for the most part, group of people who have enabled a lot of the convergence work in pulling together nanoinformatics over the years.

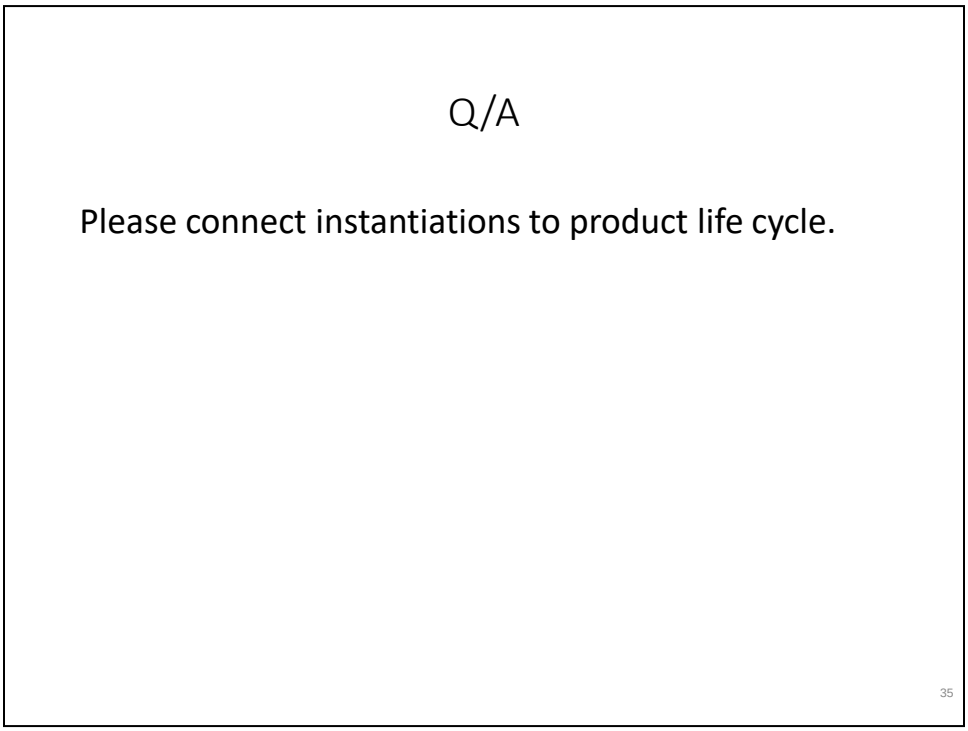
And also, for those who may or may not know, the real contribution of the National Nanotechnology Coordination Office to enabling the nanoinformatics efforts, as well. Especially because this is the last in a series of webinars, I would love to highlight Lisa Friedersdorf, Stacey Standridge, Treye Thomas, Georgios Katalagarianakis, Tom van Teunenbroek, Sally Tinkle. These are people on the U.S. and the EU side, and of course, Rhema Bjorkland, Kristin Roy, and everybody that has worked with the NNCO. (*cont.*)



>> Christine Ogilvie Hendren: (*cont.*). Without them—they are the connective tissue, being the blender, to go back to my analogy, that enabled the U.S.-EU CORs to do a lot of the work, just by way of introducing the CORs, creating a platform, creating some sort of connective root system to allow us to work together to make this stuff happen.

So while both Fred and I alluded to some volunteerism that happened, there is no way that energy could have been harnessed to go in the same direction without this type of effort. We're really appreciative. It's the long game, the investment over a long series of years, that ends up showing some fruits of this labor.

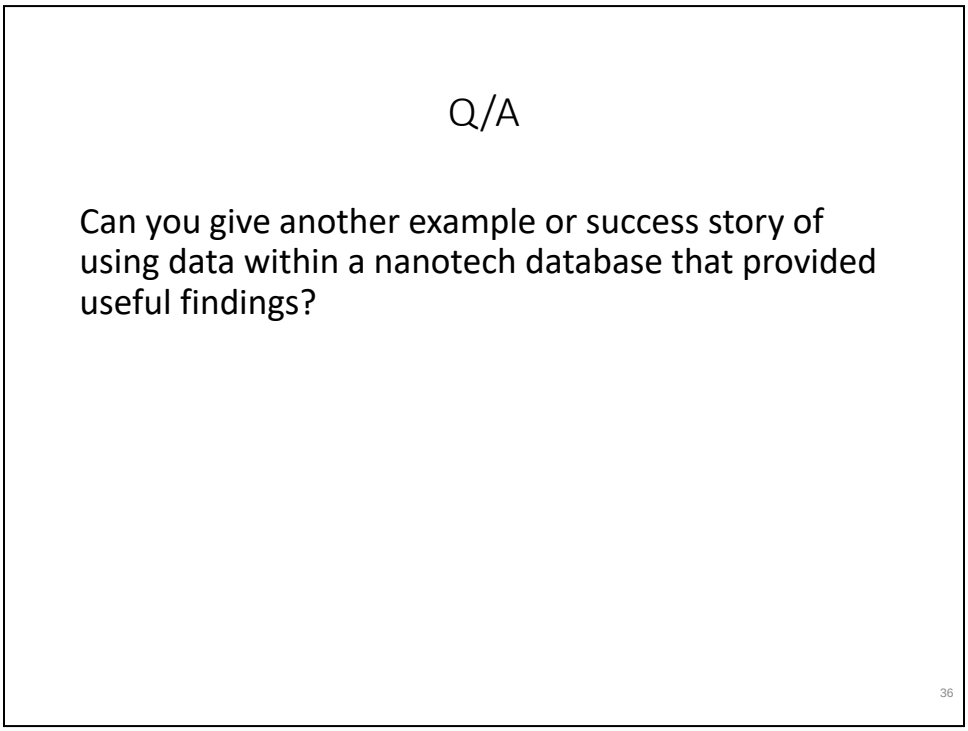
With that long acknowledgment over, I want to field some of the questions that Kristin has been kind enough to capture in this queue. I will take the first one right here and then pass one to Fred.



>> Christine Ogilvie Hendren: The question please connect instantiations to product life cycle question, I'm happy to do.

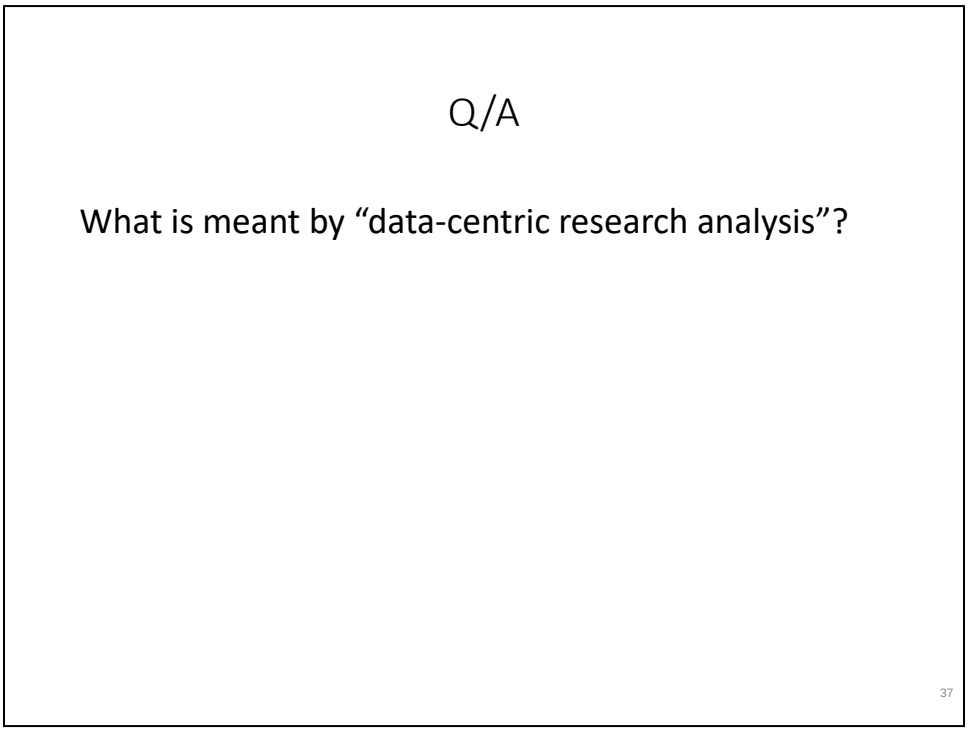
The way that I had presented it was instances of a material, and it happened to look more like over the course of an experiment. So material-as-received versus as-put-into (maybe) a medium for testing, and then maybe as prepared for characterization. You'd want to capture the surrounding media as well as the initial characteristics so that later you can back out what may have changed. Really, the system is the combined material and its surrounding media. This is the reason behind harmonizing some of the media for testing, as well. You could also, instead of taking those snapshots as slices of an experiment, you could say this is the material as it was manufactured, and then this is the same material as released into the environment or detected there later.

So these instances, as long as they include the material characterization and the surrounding media characterization, if that is done consistently, our use of the data, then, can back out how has the material changed over time as a function of its environment, and then, how can we link that to what we could forecast in terms of its impact?



>> Christine Ogilvie Hendren: So I wanted to offer, Fred, if you could answer, perhaps since you dug into a lot of great stuff with the models and the different things that came out of utilizing the datasets, once put together. Could you give another example or success story of using data within a nanotech database that provided useful findings, or (I will add a little addendum), or sets us up to do so. Would you like to take that one?

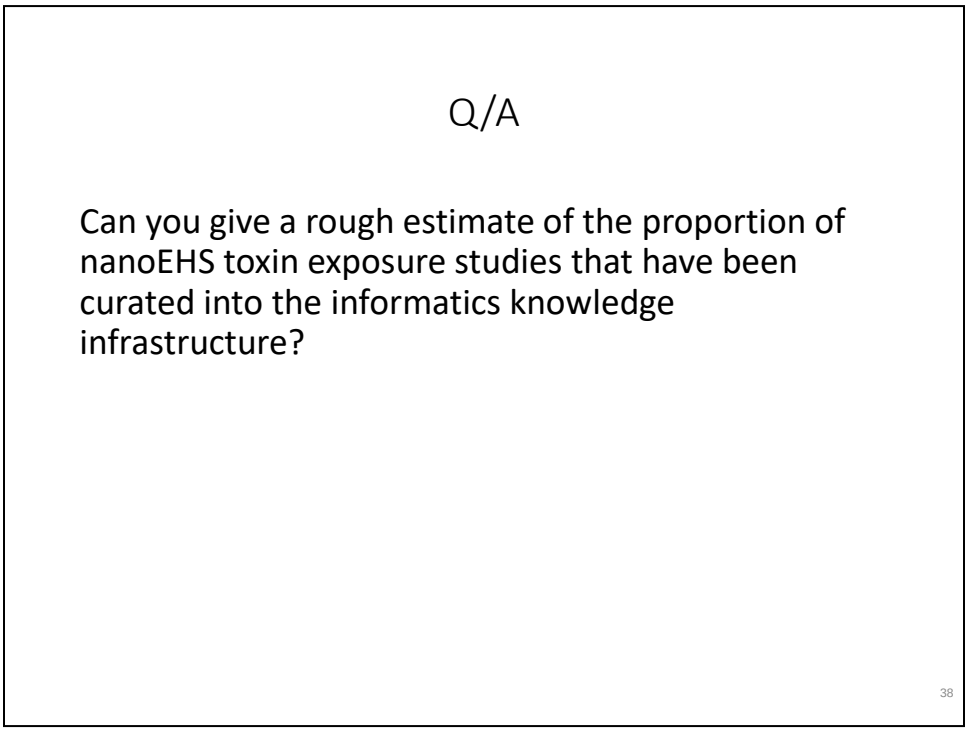
>>Fred Klaessig: Sure, Christine. Yes, one of the reasons for having the Thursday Nanotechnology Working Group session (at the NCI Center for Biomedical Informatics and Information Technology) is to provide a more practical expression of what you and I have been discussing. That is going to be one of the topic areas that Egon Willighagen and Nina Jeliazkova will be discussing. So I'd like to, shall we say, punt that to the Thursday group and allow them to have an opportunity… Especially in the area of skin sensitization, I believe, there's been a lot of activity. There's also close coordination there with the Adverse Outcome Pathways effort. So if you don't mind, I'll leave that one.



>> Fred Klaessig: The one I'd like to answer is the second question, what is meant by data-centric research analysis?

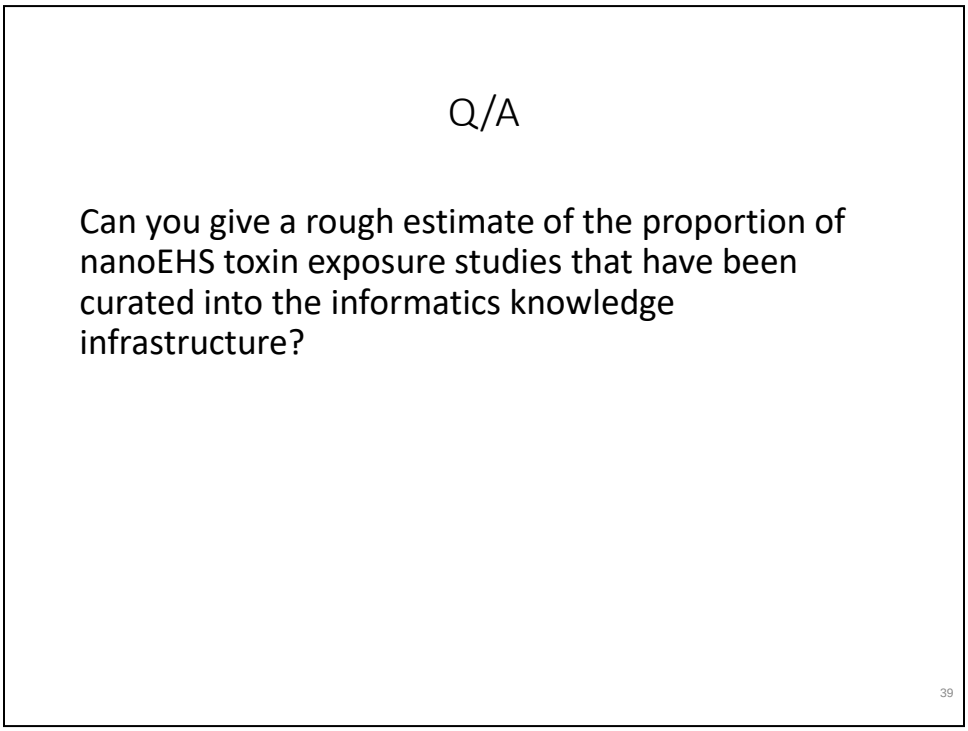
That is coming in from the bioinformatics region, and I really would want to say that it's talking to each other through the database, in the sense that a hypothesis that has been put into mathematical form and then applied to the database. What would happen if you change some of the definitions or if one came across a different grouping to pursue whether or not there's an improvement. It's not the type of thing where one writes a paper and then waits for other people to redo your work, it's more that you work through the database and connect with those people in a more immediate sense. This fits in very much with the FAIR (Findable, Accessible, Interoperable, and Reusable) effort and also the Open Science effort that's going on.

I pass the next one on the rough estimate to you, Christine.



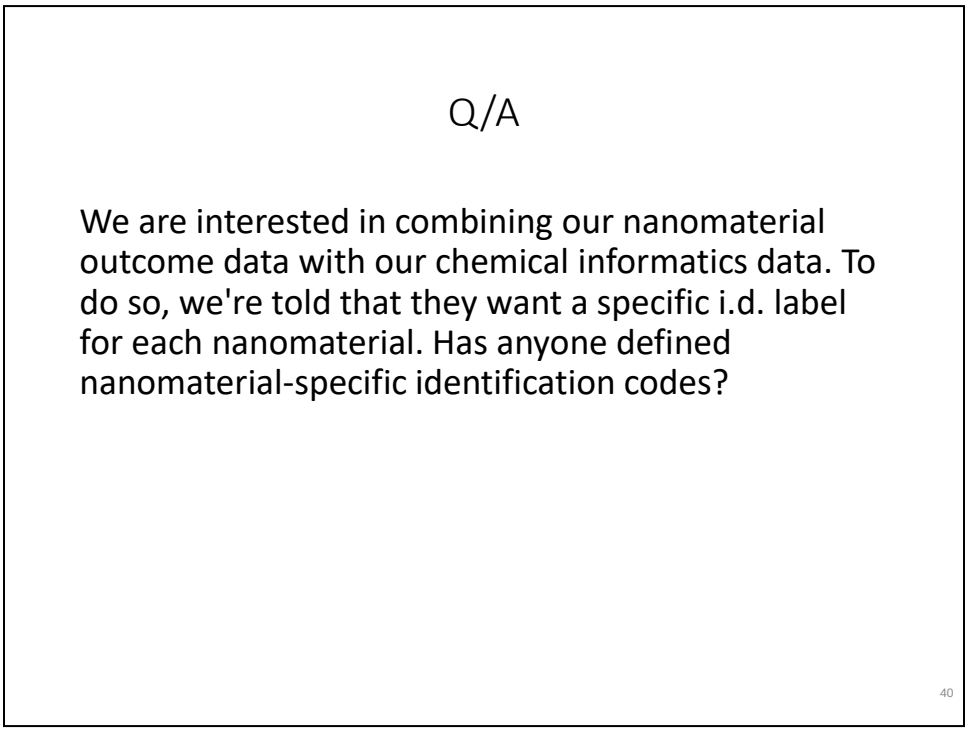
>> Christine Ogilvie Hendren: There's no way that this question can't be possibly disappointing to someone coming to this new. I would say there are different infrastructures, some of them we mentioned here, that might have different granularity of data requirements. For example, the Nanomaterial Registry loaded in characterizations of the materials. If you are talking about the NanoInformatics Knowledge Commons, which is the CEINT database that goes into the instances, so deep transformation data, we have about 200 papers there. So it's an extremely tiny portion. But since we are limited by the bottleneck of just human time and ability to process these data into the curated form, we only select those that we know will be premeditatedly interoperable—so ones with very complete metadata and ones where, pointing back to Fred's answer, we know we would be able to interrogate some of the questions. So, for example, can we actually expect a functional assay such as surface affinity to be predictive of the fate and transport of a material in a given system? We will only curate papers that we know can interrogate answers to that question. (*cont*.)

.



>> Christine Ogilvie Hendren: (*cont*.) It will remain, honestly, a small portion of studies curated into the informatics knowledge infrastructure. I believe that the best way to get critical mass going from here forward is to focus on enabling researchers to be their own curators by increasingly pulling in these parameters as we align them, and Excel-based or web-based or lab-notebook-based technology, so people who want to contribute and participate in grazing on the commons can, from the outset, add their data in, even design their experiments, based on knowing what measurements would allow them to then query on the back end.

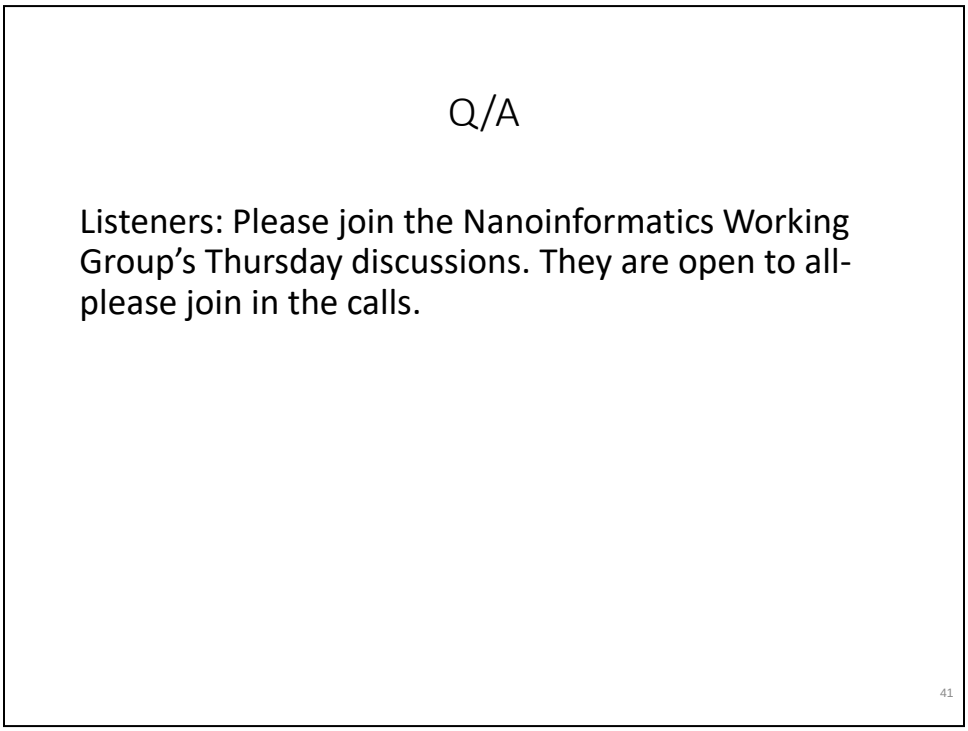
So I guess the important thing to take away is, this will remain manual and somewhat labor-intensive for the moment, in order to get the quality of data in that would allow meaningful manipulation and comparison of the data on the back end.



>> Christine Ogilvie Hendren: *Fred, would you like to choose one from the question queue?*

>> Fred Klaessig: I like this question. I'm going to betray my industrial background. I think the best way to describe fumed silica is to call it Aerosol 200 or a trade name. Very often we attempt to make, shall we say, laboratory identification of materials and commercial products be the same, when the commercial person probably has experienced a lot of quality control issues to support the uniformity of their material. That said, the colleagues in Europe are attempting to come up with an identification code, and I refer to Egon Willighagen here. That will go down to the original laboratory notebook synthesis point, so that one can follow that material throughout, shall we say, any scaleup or throughout any commercialization. I've also seen recently at the caNanoLab that they've begun putting DOI numbers onto materials, or onto datasets in fact, so one can refer back to them.

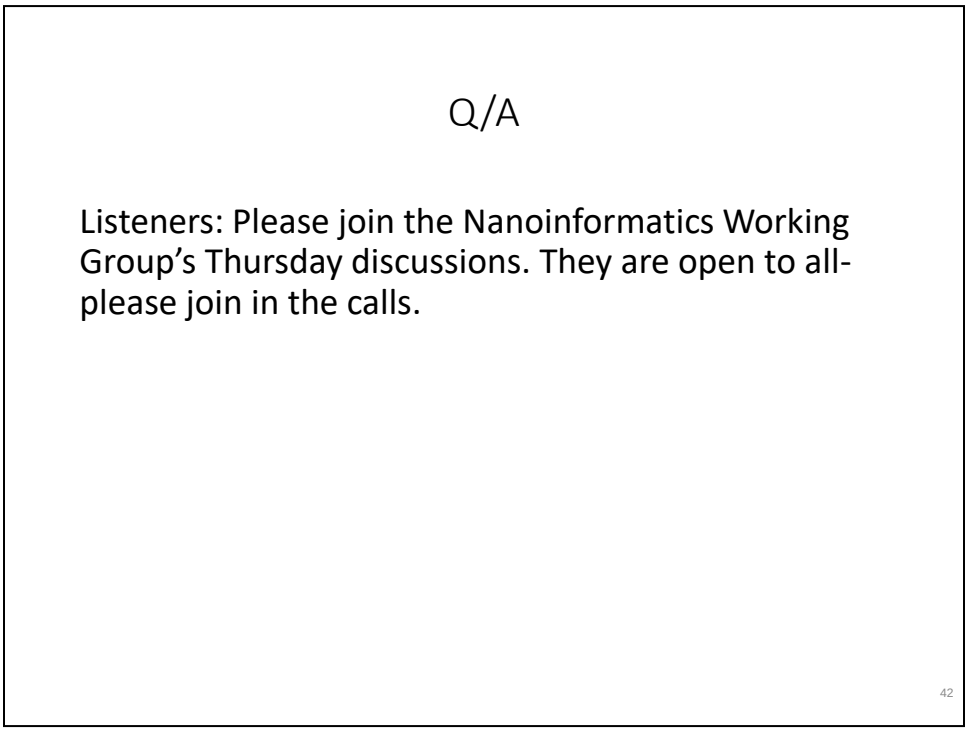
I don't know if that meets all the elements of a defined nanomaterial, since there are different definitions going around, but it would give you a, shall we say, identification code—perhaps a complex one, but would give you that, so you could refer the same material repeatedly.



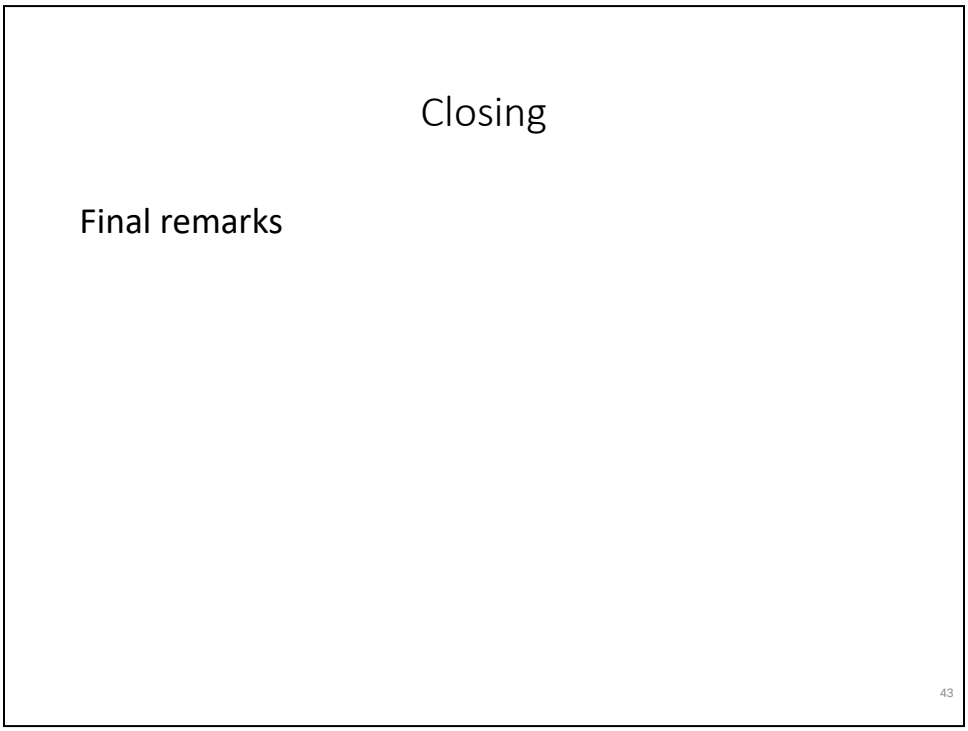
>> Christine Ogilvie Hendren: One thing to add to your last answer, Fred, is as you were speaking, I was thinking that we've mentioned the Nanoinformatics Working Group a number of times, but I don't know if at any point I made it clear that it's a completely open working group. Anybody is welcome. So there's always a mixture of things that might be useful to the community, and the community is defined by who shows up.

So if there are interests such as combining nanomaterial datasets and linking them to a chemical informatics dashboard, for example, I would encourage anybody on the call with interests in this type of effort, or with ideas or questions, to consider calling in to that group and pose those questions to the group.

It's currently led by Iseult Lynch, who's the Director of NanoCommons and a number of other efforts in the EU, and Luisa Russell, who Fred mentioned earlier is doing excellent work with caNanoLab and many other things. So just a call to anyone who wants to continue this conversation beyond what we're able to cover in today's webinar, to say it is a self-assembled group of similarly motivated people with a bunch of diverse ideas. You'd be welcome to call in. I will risk inefficiency by saying you would be welcome to email me. You can Google the NCIP Nano Working Group, and if you can't find that, you can email me at christine.hendren@duke.edu.



>>Fred Klaessig: To add to that, the Thursday meeting of the group is held 11:00 a.m. ET (Nano WG 2020 calendar), There are two practical discussions on using databases in this topic area, I think I gave contact information for Mervi Heiskanen, and Rhema provided a live link in her invitations to everyone. So everyone should feel free to join.



>> Christine Ogilvie Hendren: All right. Well, we have two minutes left. It does appear that we don't have any more questions in the queue. If anybody would like to add another one in, we could answer it with these last few minutes. And otherwise, I'll just wait for one moment and see. While we wait, the person who asked, “Can you give another example or success story of a nanotech database that produced useful findings,” I guess I would take the opportunity to say, quite honestly, I think the best success stories are in our future, only because so much of the groundwork has been laying the infrastructure to start to get people to add in their data.

I think one place, and Fred made a really good point in this during his part of the presentation, one place we will see a lot of the benefits is in the combination of data and modeling. So for example, there's a nanoproduct hazard and exposure assessment tool that's been developed for the Consumer Product Safety Commission that allows us to look at toxicity data from the literature, construct dose-response curves, and then through experiment and kind of user assumption entry, to then project estimates on what you would expect to come out of nano-enabled products when they are released in the environment, and how those exposures and those concentrations might compare to the dose-response efforts. So creative matchups between data and estimates—I think that we're starting to see that type of success. But there's a lot more to come as well.

We have zero more minutes left. I want to thank you, Fred, for being willing to co-present with me. Thanks, Rhema and Kristin, for hosting us today. And thank you, everyone, for your time in tuning in to our webinar.