NAN MATERIAL REGISTRY

Alexander Tropsha, Kimberly Guzan RTI International December 2, 2013

FEDERAL NANOTECHNOLOGY INITIATIVE



The NNI Coordinates R&D in nanoscale science, engineering, and technology



http://nano.gov/sites/default/files/pub resource/nki nsi white paper - final for web.pdf

NANOMATERIALREGISTRY

MINIMAL INFORMATION CHARACTERIZATION DATA



Definition of Minimal Information is common for different disciplines

Journal home Advance online

publication Current issue

Archive

Web Focuses

Article Series

Multimedia

Posters

Journal information

⊠ Feedback for editors

NPG services

Help Authors and Referees Librarian gateway Advertising information work@npg Reprints and permissions

NPG resources

SciBX

Nature Reviews Drug Discovery 10, 661-669 (September 2011) | doi:10.1038/nrd3503

OPINION

Minimum information about a bioactive entity (MIABE) See also: <u>Correspondence by Lentini</u> | <u>Author Reply by Orchard et al.</u>

Sandra Orchard¹, Bissan Al-Lazikani², Steve Bryant³, Dominic Clark¹, Elizabeth Calder⁴, Ian Dix⁴, Ola Engkvist⁵, Mark Forster⁵, Anna Gaulton¹, Michael Gilson⁷, Robert Glen⁸, Martin Grigorov⁹, Kim Hammond-Kosack¹⁰, Lee Harland^{11,12}, Andrew Hopkins¹³, Christopher Larminie¹⁴, Nick Lynch⁴, Romeena K. Mann¹⁴, Peter Murray-Rust⁸, Elena Lo Piparo⁹, Christopher Southan¹⁵, Christoph Steinbeck¹, David Wishart¹⁶, Henning Hermjakob¹, John Overington¹ & Janet Thornton¹ <u>About the authors</u>

top 1

Bioactive molecules such as drugs, pesticides and food additives are produced in large numbers by many commercial and academic groups around the world. Enormous quantities of data are generated on the biological properties and quality of these molecules. Access to such data - both on licensed and commercially available compounds, and also on those that fail during development - is crucial for understanding how improved molecules could be developed. For example, computational analysis of aggregated data on molecules that are investigated in drug discovery programmes has led to a greater understanding of the properties of successful drugs. However, the information required to perform these analyses is rarely published, and when it is made available it is often missing crucial data or is in a format that is inappropriate for efficient data-mining. Here, we propose a solution: the definition of reporting guidelines for bioactive entities – the Minimum Information About a Bioactive Entity (MIABE) – which has been developed by representatives of pharmaceutical companies, data resource providers and academic groups.

MINIMAL INFORMATION CHARACTERIZATION DATA



Composition



Size



Size Distribution



Shape



Aggregation/ Agglomeration State



Surface Area



Surface Charge



Surface Chemistry



Surface Reactivity



Purity



Solubility



Stability

NAN MATERIAL REGISTRY

MAIN PCC DATA



Minimal Information About Nanomaterials for Physico-Chemical Characteristics



Minimal Information = PCC data + Metadata

NANOMATERIALREGISTRY

DATA CURATION WORKFLOW





NAN OMATERIAL REGISTRY

Data curation is critical to achieve high quality databases and enable modeling studies **ChemBark**

News, Analysis, and Commentary for the World of Chemistry & Chemical Research

« Hacks for Septa

Organometallics Responds to the Dorta Situation »

A Disturbing Note in a Recent SI File

August 6th, 2013

A recently published ASAP <u>article</u> in the journal *Organometallics* is sure to raise some eyebrows in the chemical community. While the paper itself is a straightforward study of palladium and platinum bis-sulfoxide complexes, page 12 of the corresponding Supporting Information <u>file</u> contains what appears to be an editorial note that was inadvertently left in the published document:

Emma, please insert NMR data here! where are they? and for this compound, just make up an elemental analysis...



66

NS

as

of

ed of

ited

re is

se

This statement goes beyond a simple embarrassing failure to properly edit the manuscript, as it appears the first author is being instructed to fabricate data. Elemental analyses would be very easy to fabricate, and

long-time readers of this blog will recall how fake elemental analyses were pivotal to Bengu Sezen's <u>campaign of fraud</u> in the work she published from 2002 to 2005 out of Dalibor Sames' lab at Columbia.

The compound labeled **14** (an acac complex) in the main paper does not appear to correspond to compound **14** in the SI. In fact, the bridged-dichloride compound appears to be listed an as unlabeled intermediate in Scheme 5, which should raise more eyebrows. Did the authors unlist the compound in order to avoid having to provide robust characterization for it?

ChemBark is contacting the <u>corresponding author</u> for comment, and his response will be posted in full when we receive it.

a 8 (12

Fl

Di

B

re

d

Full Pape

Are th

Douglas Y

^a US Envire
E-mail: yc
^b Pegasus T

Keywords: I relationships

Received: Ju

DOI: 10.100



nanomaterial record



A **DATA CURATION TOOL** facilitates the progression of nanomaterial entries through the curation process to the Nanomaterial Registry website



//www.nanomaterialregistry.org

ToxCa... 🗅 file:///C:/Users/atro...) Google 🕒 New Tab 🕒 Save to Mendeley 📰 Savings & Investme...

wtharvey.com/m8n...

🥑 Сериал Лист ожида...



THE REGISTRY RECORDS

1284 Nanomaterials as of 10/31/2013 Visualize Data - Click Here



A TOOL FOR THE NANOMATERIAL COMMUNITY

The large quantity of nanomaterial-focused research and production has created a vast and growing body of nanomaterial-centric data. Yet, critical knowledge gaps in nanotechnology need to be addressed so that the use of nanomaterials is not outpacing the understanding of their implications. The Nanomaterial Registry is a data-driven tool aimed at enabling researchers to close this knowledge gap. As researchers interact with the information in this central data repository, knowledge will be extracted and used to guide new research and, ultimately, the safe use of nanomaterials. The Registry is set up to provide researchers with a convenient data management and sharing plan. Both NIH and NSF require grant recipients to have a data management plan. Our curation team is ready to accept your public-ready data, archive them, integrate them with the larger data repository, and share them publicly for you. Visit our Data Management Plan page to download boilerplate text for your proposal or contact us for customized text!

To access the data in the Nanomaterial Registry, search or browse the database using the buttons on this Home page. From a guery results table, you can request detailed information on specific nanomaterial records. compare multiple nanomaterials in side-by-side view, and export data to your desktop in an easy-to-analyze Excel spreadsheet. For additional guidance, please refer to our top-menu pages, including the User Guide. To ask questions or provide feedback, please refer to our Contact Us page.

Thank you for using the Nanomaterial Registry!

LATEST NEWS

October 2013 - A new principal investigator and coprincipal investigator have joined the Nanomaterial Registry team! Please join us ... Read more

October 2013 ·

Nanoinformatics 2013: Informatics for Nanomanufacturing will be held October 15 in Philadelphia, PA. This year's workshop ... Read more

SEE ALL NEWS >

REGISTRY DATA

User Guide

Glossary

News

Registry Data

Standards and Guidance

Characterization Profile for Nanomaterials



	AGGREGATION / AGGLOMERATION STATE	168	70	98		
	COMPOSITION	1175	945	230		
	PARTICLE SIZE	834	330	322	174	8
	PURITY	39	6	33		
	SHAPE	369	343	3	23	
	SIZE DISTRIBUTION	119	112	3		4
	SOLUBILITY	187	38	149		
	STABILITY	101	17	28	56	
	SURFACE AREA	99	90	9		
	SURFACE CHARGE	328	296	2	30	
	SURFACE CHEMISTRY	163	162	1		
	SURFACE REACTIVITY	6	6			



The <u>DATA</u> and data analytics interplay: *bioinformatics example*

Google labs Books Ngram Viewer



Search in Google Books:

The <u>DATA</u> and data analytics interplay : *cheminformatics example*

Done



😔 Internet | Protected Mode: C

The <u>DATA</u> and data analytics interplay: *nanoinformatics example*

books Ngram Viewer



Major challenge: speed up and streamline data collection to grow NR

Current data publication workflow

- Primary data collection
- <u>Electronic</u> laboratory data record (maybe, ELNs?)
- Data formatting for publication
- Submission, review, publication
- PDF of published paper

Current data entry into the Registry

- PDF of published paper
- Manual data extraction and conversion to <u>electronic</u> format
- Data curation
- Manual data entry into Nanomaterial Registry
- Data made available through NR portal

Primary data generation is funded by federal grants

NIH	Research	earch Portfolio Online Reporting Tools			Search NANOMATERIALS Q								
	1	/		НО	ME ABOU	JT RePORT	FAQs	GLOSSARY C	ONTACT U				
QUICK LINKS	RES	EARCH ORGANIZATION	NS WORKFO	DRCE FU	NDING	RE	PORTS	LINKS &	DATA				
Home > <u>RePORTER</u> > S	Search Results			M	RePOR	TER La	ogin Regis	ter System Healt	th: 📕 GREEN				
Search Results Save Query Form Save Query Save Query													
							Exp	All Projects	GO				
PROJECTS ? P	UBLICATIONS	PATENTS DATA & VISUALIZE MAP L	NKS 🗗 NEWS & MORE	ন্দ্র									
There were 226 resu	ilts matching you	r search criteria.	Records per page 25				S	how/Hide Search (Criteria 🛹				
Click on the column h	eader to sort the	results	12348910				Page	1 of 10 Next La	st 🕨 🕨				
T: Application Type; Act: Activity Code; Project: Admin IC, Serial No.; Year: Support Year/Supplement/Amendment													
T Act Projec	t Year Sub#	Project Title	Contact PI/ Project Leader	Organization	FY	Admin IC	Funding IC	FY Total Cost by IC	Similar Projects				
<u>5 U54 CA1518</u>	<u>81</u> 0 <u>4</u> 5131	EDUCATION/TRAINING AND OUTREACH	AMUI, MANSOOR M.	NORTHEASTERN UNIVERSITY	2013	NCI		\$51,921					
5 U54 CA1516	<u>62 04</u> 8773	ADMINSTRATIVE CORE	BAKER, IAN	DARTMOUTH COLLEC	GE 2013	NCI		\$251,397					
5 U19 ES01954	4 <u>5 04</u> 5579	CORE 1 - NANOMATERIALS SYNTHESIS FABRICATION, AND CHARACTERIZATION	BANEYX, FRANCOIS C	UNIVERSITY OF WASHINGTON	2013	NIEHS		\$372,124					
5 R25 CA1540	<u>15 04</u>	TRAINING THE NEXT GENERATION OF RESEARCHERS IN CANCER NANOTECHNOLOGY AT THE NCI M	BASHIR, RASHID et al.	UNIVERSITY OF ILLIN URBANA-CHAMPAIG	0IS 2013	NCI	NCI	\$344,911					
			RA74L01/A										

Important public repositories rely on data uploading by users





ADIT depositors can submit their sequences to serve as puzzle challenges for the Foldit community. For more information, see here.



f you are depositing a protein-protein, protein-DNA, protein-RNA or protein-peptide complex, you may consider submitting your structure as a target to CAPRI (Critical Assessment of Predicted Interactions) after depositing your entry to the PDB.

To start a new ADIT session, select the experimental method and the molecular structure type. Then press the BEGIN button.

Major challenge: speed up and streamline data collection to grow NR

Current data publication workflow

- Primary data collection
- <u>Electronic</u> laboratory data record (maybe, ELNs?)
- Data formatting for publication
- Submission, review, publication
- PDF of published paper

Current data entry into the Registry

- PDF of published paper
- Manual data extraction and conversion to <u>electronic</u> format
- Data curation
- Manual data entry into Nanoregistry
- data made available through Nanoregistry portal

Enabling mechanisms (COR effort!)

- Funding agencies: require that that funded projects upload their data on PCCs and other tested properties of ENMs to Nanomaterial Registry as part of data sharing plan
- Journals: require data upload as a condition for publications
- Nanomaterial Registry: provide user-friendly tools facilitating data upload/curation (cf. PDB).

NAN MATERIAL REGISTRY

THANK YOU!

www.nanomaterialregistry.org nanoregistry@rti.org

