



NanoCommons

Nano-Knowledge Community

Spreadsheets in nanosafety research

***Egon Willighagen, orcid:0000-0001-7542-0286, @egonwillighagen
NanoCommons Workshop, 2020-04-28, #nanocommons***

***This project has received funding from the European Union Horizon 2020
Programma (H2020) under grant agreement no. 731032.***

CC-BY 4.0 International (unless otherwise specified)

What is a spreadsheet?

- Wikipedia:
 - *“A spreadsheet is a computer application for organization, analysis and storage of data in tabular form.”*
- National Cancer Institute Thesaurus (NCIT ontology):
 - *“A document that organizes data into rows and columns.”*
 - <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C85438>
- Tools:
 - Microsoft Excel, Google Spreadsheet, LibreOffice Calc
 - Comma-Separated Values (CSV), Tab-Separated Values (TSV)

Supplementary Information

Meta-Analysis of Nanoparticle Cytotoxicity via Data-Mining the Literature

Hagar I. Labouta*, Nasimeh Asgarian, Kristina Rinker and David T. Cramb*

✓ **Cite this:** *ACS Nano* 2019, 13, 2, 1583-1594

Publication Date: January 28, 2019 ✓

<https://doi.org/10.1021/acsnano.8b07562>

Copyright © 2019 American Chemical Society

[RIGHTS & PERMISSIONS](#)

Article Views

935

Altmetric

4

Citations

3

[LEARN ABOUT THESE METRICS](#)



PDF (3 MB)



Loading Institutional Login Options...

More Access Options



Supporting Info

SUBJECTS: [Coating materials](#), [Assays](#), [Toxicity Nanoparticles](#)

Abstract

Developing predictive modeling frameworks of potential cytotoxicity of engineered nanoparticles is critical for environmental and health risk analysis. The complexity and the heterogeneity of available data on potential risks of nanoparticles, in addition to interdependency of relevant influential attributes, makes it challenging to develop a generalization of nanoparticle toxicity behavior. Lack of

Supplementary Information

nn8b07562_si_001.xlsx - LibreOffice Calc

File Edit View Insert Format Styles Sheet Data Tools Window Help

Calibri 11 B I U T A % 00 00 00


B57 f_x Σ = I

	A	B	C	D	E	F	G	H	I	J	K
1	Nanoparticle	Type: Organic (O)/inorganic (I)	coat	Diameter (nm)	Concentration μM	Zeta potential (mV)	Cells	Cell line (L)/primary cells (P)	Human(H)/Animal(A) cells	Animal?	Cell morphology
2	CeO2	I		20.3	0.0024778287664		L929	L	A	Mouse	Fibroblast
3	CeO2	I		20.3	0.0049556575328		L929	L	A	Mouse	Fibroblast
4	CeO2	I		20.3	0.0099113150655		L929	L	A	Mouse	Fibroblast
5	CeO2	I		20.3	0.019822630131		L929	L	A	Mouse	Fibroblast
6	CeO2	I		20.3	0.0396452602621		L929	L	A	Mouse	Fibroblast
7	Se	I		79.6	1.307545634E-06	0	PC3	L	H		Epithelial
8	Se	I		79.6	2.615091268E-06	0	PC3	L	H		Epithelial
9	Se	I		79.6	5.230182535E-06	0	PC3	L	H		Epithelial
10	Se	I		79.6	7.845273803E-06	0	PC3	L	H		Epithelial
11	CuO	I		52.51	3.471326159E-05	-39.67	A549	L	H		Epithelial
12	CuO	I		52.51	8.678315397E-05	-39.67	A549	L	H		Epithelial
13	CuO	I		52.51	0.0001735663079	-39.67	A549	L	H		Epithelial
14	CuO	I		52.51	3.471326159E-05	-39.67	A549	L	H		Epithelial
15	CuO	I		52.51	8.678315397E-05	-39.67	A549	L	H		Epithelial
16	CuO	I		52.51	0.0001735663079	-39.67	A549	L	H		Epithelial
17	CuO	I		239	3.681525162E-11		HCMC	P	H		Endothelial
18	CuO	I		239	3.681525162E-10		HCMC	P	H		Endothelial
19	CuO	I		239	3.681525162E-09		HCMC	P	H		Endothelial
20	CuO	I		239	3.681525162E-08		HCMC	P	H		Endothelial
21	CuO	I		239	1.840762581E-07		HCMC	P	H		Endothelial
22	CuO	I		239	3.681525162E-07		HCMC	P	H		Endothelial
23	CuO	I		239	7.363050324E-07		HCMC	P	H		Endothelial
24	CuO	I		239	1.840762581E-06		HCMC	P	H		Endothelial
25	CuO	I		239	3.681525162E-06		HCMC	P	H		Endothelial

Mistake #1: Excel dates

Correspondence | [Open Access](#) | Published: 23 June 2004

Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics

[Barry R Zeeberg](#), [Joseph Riss](#), [David W Kane](#), [Kimberly J Bussey](#), [Edward Uchio](#), [W Marston Linehan](#), [J Carl Barrett](#) & [John N Weinstein](#) 

BMC Bioinformatics 5, Article number: 80 (2004) | [Cite this article](#)

114k Accesses | 44 Citations | 514 Altmetric | [Metrics](#)

Abstract

Background

When processing microarray data sets, we recently noticed that some gene names were being changed inadvertently to non-gene names.

Results

A little detective work traced the problem to default date format conversions and floating-point format conversions in the very useful Excel program package. The date conversions affect at least 30 gene names; the floating-point conversions affect at least 2,000 if Riken identifiers are included. These conversions are irreversible: the original gene names cannot be



Mistake #1: Excel dates

	<i>gene names</i>	<i>internal date format</i>	<i>default date format</i>		<i>gene names</i>	<i>internal date format</i>	<i>default date format</i>		<i>gene names</i>	<i>internal date format</i>	<i>default date format</i>
1	APR-1	35885	1-Apr		OCT-1	36068	1-Oct		SEP2	36039	2-Sep
2	APR-2	35886	2-Apr		OCT-2	36069	2-Oct		SEP3	36040	3-Sep
3	APR-3	35887	3-Apr		OCT-3	36070	3-Oct		SEP4	36041	4-Sep
4	APR-4	35888	4-Apr		OCT-4	36071	4-Oct		SEP5	36042	5-Sep
5	APR-5	35889	5-Apr		OCT-6	36073	6-Oct		SEP6	36043	6-Sep
6	DEC-1	36129	1-Dec		OCT1	36068	1-Oct		SEPT1	36038	1-Sep
7	DEC-2	36130	2-Dec		OCT11	36078	11-Oct		SEPT2	36039	2-Sep
8	DEC1	36129	1-Dec		OCT2	36069	2-Oct		SEPT3	36040	3-Sep
9	DEC2	36130	2-Dec		OCT3	36070	3-Oct		SEPT4	36041	4-Sep
10	MAR1	35854	1-Mar		OCT4	36071	4-Oct		SEPT5	36042	5-Sep
11	MAR2	35855	2-Mar		OCT6	36073	6-Oct		SEPT6	36043	6-Sep
12	MAR3	35856	3-Mar		OCT7	36074	7-Oct		SEPT7	36044	7-Sep
13	NOV1	36099	1-Nov		SEP-1	36038	1-Sep		SEPT8	36045	8-Sep
14	NOV2	36100	2-Nov		SEP-2	36039	2-Sep		SEPT9	36046	9-Sep
15					SEP1	36038	1-Sep				



Mistake #1: Excel dates

ncbi_mgd Human Chromosome 2, Mouse Homology R...

Address: http://www.ncbi.nlm.nih.gov/Homology/view.cgi?map=ncbi_mgd&chr=2&tax_id=9606

NCBI Human-Mouse Homology Map

Map: ncbi vs. mgd Master: Human Chromosome: 2 Go

View as text

Human STS	Cytogen Pos	Human Symbol	Mouse chr	Mouse Symbol	cM Position	Mouse STS
	2p25.3	DKFZP586F1318	12	Sh3y11		
	2p25	ACPI	12	Acp1		
	2p25	TPO	12	Tpo	15	
	2p25.3	MYT1L	12	Myt1l	14	
	2p25.3	MGC3279	12	1010001H16Rik		
	2q32.1-q36.3	ALLC	12	Alle		
	2p25	SOX11	12	Sox11		
	2p25.3	KIAA01	12	Ubcap4-pending		
	2p25	ID2	12	Id2		
	2q37	ASB1	1	Asb1		
	2q35-q37	GPC1	1	Gpc1		
	2q37	ATSV	1	Kif1a		
	2q37.3	GPR35	1	Gpr35		
	2q37.3	CAPN10	1	Capn10		
	2q37.3	PPP1R7	1	Ppp1r7		
	2q37	HDLBP	1	Hdlbp	55.3	
	2q37	NEDD5	1	2-Sep		
	2q37.3	STK25	1	Stk25	58	
	2q36-q37	COL4A3 *	1	Col4a3		
	2q35-q37	GPC1 *	1	Gpc1		
	2q37.3	GPR35 *	1	Gpr35		
	2q37.3	PDCD1 *	1	Pded1		
	2q37	UGT1A6 *	1	Ugt1a6		
	2q37.3	HES6 *	1	Hes6		
	2q37	SLC19A3 *	1	Slc19a3	51	
	2q37	SLC19A3 *	1	Slc1a1	51	

Mistake #2: not enough columns

- Requires additional processing, but okay for humans
- Call values include units with different units
- Different units
- Average, standard deviation, ranges

4
Reference
doi: 10.1021/nn406018q
doi: 10.1021/nn406018q

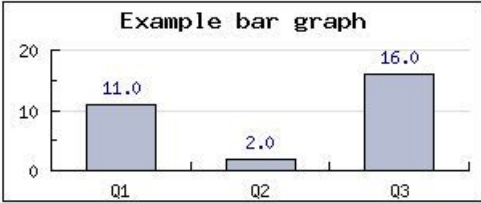
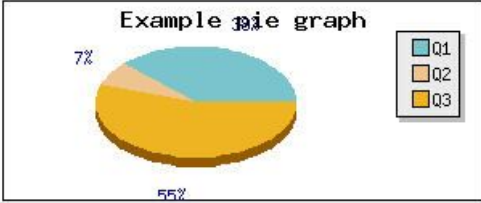
Better:

10	11	12	13	14
Endpoint	Unit	Qualifier	value	error
Net cell association	mL/ug(Mg)	Mean	0.02751	0.01654
Net cell association	mL/ug(Mg)	Mean	0.49705	0.08013



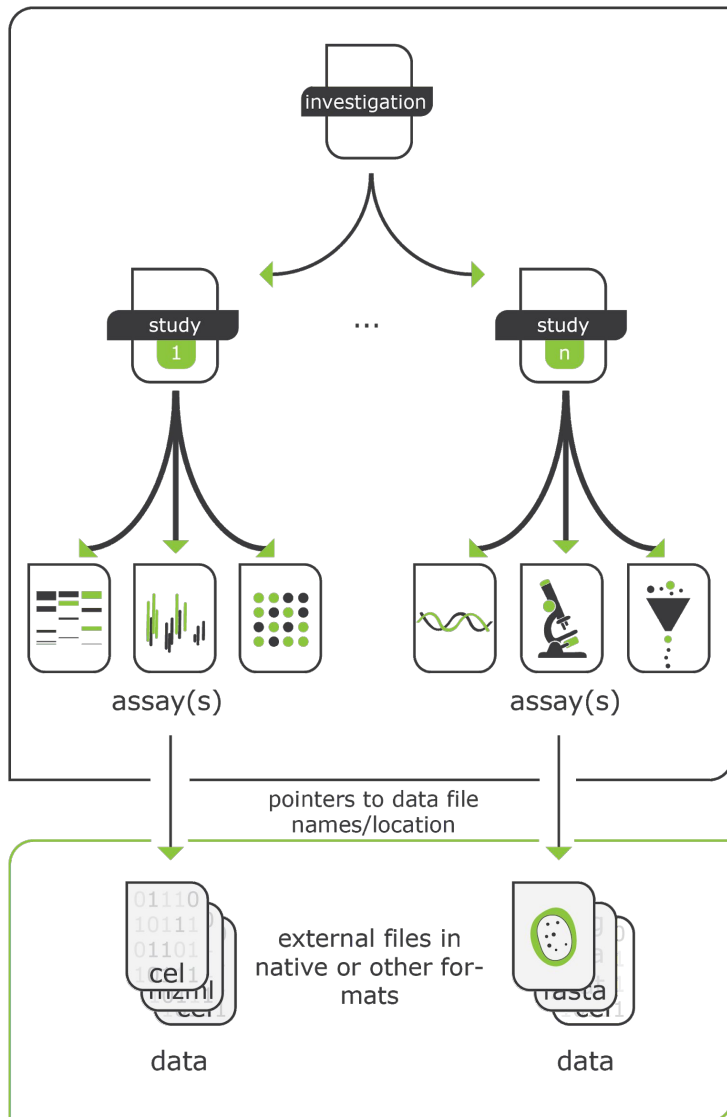
Mistake #3: not rectangular

Save - Home - << - >> - Up - Down - 13 x 30 - A1 - =C3 font-weight:bold; Save X - Formulas - Help

	Group1		Group2		Group3			
	A	Col1 - B	Col2 - C	Col3 - D	E	F	G	H
Group#1 1	test2				abc http://www.cnn.com test http://www.heise.de			
Row1 2		44.32%	-2,000,001.00 €	\$10.00	http://www.cnn.com			
Row2 3		test3	test2	20	thomas.bley@simple-groupware.de			
Row3 4				30		test2C^3+C3	test2C^3-C3	C^3-C3test2C^3
Group#2 5		locked			undefined	75		
Row1 6						60		
Row2 7								<"e'st100te's"t>
Row3 8		1	2	3				
9		4	5	6				
10								200
11		sum(B8:D8)	6			Example bar graph	Example pie graph	
12		min(B8:D8)	1			Q1	11	
13		max(B8:D8)	3			Q2	2	
14		avg(B8:D8)	2			Q3	16	
15		count(B8:D8)	3					

Insert Row - Insert Column - Delete Row - Delete Column Copyright (C) 2006 by Thomas Bley. Group1 - Group#1 long name - A1

ISA-Tab



isa

investigation

high level concept to link related studies

study

the central unit, containing information on the subject under study, its characteristics and any treatments applied.

*a study has associated **assays***

assay

test performed either on material taken from the subject or on the whole initial subject, which produce qualitative or quantitative measurements (data)

ISA-Tab-Nano

	A	B	D	E	F	G	H	I
1	Material Source Name	Material Name	Material Description	Material Synthesis	Material Design Rationale	Material Intended Application	Term Accession Number	Term Source REF
2	LAB-1	g4.5_carboxyl-terminated dendrimer	G 4.5 carboxyl-terminated dendrimer			delivery of image contrast agent		
3	LAB-1	g4.5_core						
4	LAB-1	g4.5_branch						
5	LAB-1	g4.5_carboxyl_group						

	J	K	L	M	N	O
1	Material Type	Term Accession Number	Term Source REF	Material Chemical Name	Term Accession Number	Term Source REF
2	nanoparticle sample; dendrimer	NPO_1404; NPO_735	NPO;NPO			
3	dendrimer core molecule	NPO_803	NPO	diaminobutane		
4	dendrimer branch	NPO_776	NPO	poly(amidoamine)	53199	CHEBI
5	dendrimer terminal group	NPO_695	NPO	carboxyl group	46883	CHEBI

	P	Q	R	S	T
1	Characteristics [molecular weight {NPO:NPO_1171}]	Unit	Term Accession Number	Term Source REF	Characteristics[theoretical number of groups]
2	27	kDa	UO_0000222	UO	
3					
4					
5					128

JRC Harmonized Templates



EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F – Health, Consumers & Reference Materials
Consumer Products Safety Unit

NANoREG data logging templates for the environmental, health and safety assessment of nanomaterials

Data is a critical matter in the nanomaterials scientific community. Indeed, up to now, the recording (logging) of experimental data produced in several projects has been insufficiently harmonised. Data (and knowledge) often remains linked to *ad hoc* file systems or templates generated for the only project they serve and, mostly, just for its duration. In the recent dialogue within the EU NanoSafey Cluster¹, it became clear how important the harmonisation of data logging it is to obtain fully exploitable data. This allows, among others, a better sharing, comparability and analysis.

One promising data storage and sharing logic is ISA-TAB-Nano². It is a powerful way of structuring and storing data. It relates data to protocols, standard operating procedures (SOPs) or methods that have been used to generate it. Yet, low user-friendliness limits the applicability of ISA-TAB-Nano in a "laboratory environment". Hence, the EU-funded FP7 project NANoREG³ has created a set of Excel[®] [templates](#) for use by researchers in the field. They are easily convertible to the ISA format with, for instance, spreadsheet parsers developed by the EU FP7 project eNanoMapper⁴. This suite of Excel[®] files is the result of a collective effort by many involved project partners until 30 September 2016.

The [downloadable zipped archive](#) contains three different zipped folders, one for each "module": physicochemical characterisation, mammalian toxicology *in vitro* and mammalian toxicology *in vivo*

GRACIOUS extension



EUROPEAN COMMISSION
DIRECTORATE-GENERAL JOINT RESEARCH CENTRE
Directorate F – Health, Consumers & Reference Materials
Consumer Products Safety Unit

GRACIOUS data logging templates for the environmental, health and safety assessment of nanomaterials

The harmonised recording of experimental data on nanomaterial properties generated in different research projects is a key issue in nanosafety. In 2017, NANoREG¹ published a set of Excel[®] templates² aimed to facilitate the reporting of data on 33 endpoints (including 12 physicochemical properties, 9 *in vitro* mammalian toxicity assays and 12 *in vivo* mammalian toxicity assays) generated during laboratory work. Each template is based on ISA-TAB-Nano^{3,4} and relates the result of the measurement to the experimental conditions, protocols, method and instrument that have been used to generate it, thus ensuring reproducibility, comparability and re-use of the data by other scientists. More recently, GRACIOUS⁵ has decided to use some of these templates to store experimental data generated in the project and to create new templates with the same structure for physicochemical properties relevant for the project but not covered by existing NANoREG templates.



doi: 10.2760/142959



But what about mistake #4?

	A	B	D	E	F	G	H	I
1	Material Source Name	Material Name	Material Description	Material Synthesis	Material Design Rationale	Material Intended Application	Term Accession Number	Term Source REF
2	LAB-1	g4.5_carboxyl-terminated dendrimer	G 4.5 carboxyl-terminated dendrimer			delivery of image contrast agent		
3	LAB-1	g4.5_core						
4	LAB-1	g4.5_branch						
5	LAB-1	g4.5_carboxyl_group						

	J	K	L	M	N	O
1	Material Type	Term Accession Number	Term Source REF	Material Chemical Name	Term Accession Number	Term Source REF
2	nanoparticle sample; dendrimer	NPO_1404; NPO_735	NPO;NPO			
3	dendrimer core molecule	NPO_803	NPO	diaminobutane		
4	dendrimer branch	NPO_776	NPO	poly(amidoamine)	53199	CHEBI
5	dendrimer terminal group	NPO_695	NPO	carboxyl group	46883	CHEBI

	P	Q	R	S	T
1	Characteristics [molecular weight {NPO:NPO_1171}]	Unit	Term Accession Number	Term Source REF	Characteristics[theoretical number of groups]
2	27	kDa	UO_0000222	UO	
3					
4					
5					128

Ontology Annotation

A concept in an ontology has at least:

1. A label / name
 - a. E.g. *“titanium dioxide nanoparticle”*
2. A description
 - a. E.g. *“A nanoparticle consisting of titanium dioxide.”*
3. A parent class
 - a. E.g. *“inorganic nanoparticle”*
4. An Internationalized Resource Identifier (IRI)
 - a. E.g. *http://purl.obolibrary.org/obo/CHEBI_51050*

Ontologies and Spreadsheets: nmdataparser

Examples

Converting NanoSafety Cluster Excel spreadsheets

- Example `xlsx` file `INVITRO_VIABILITY_Trypanblue_TEST.xlsx`
- Example `json` configuration file `INVITRO_VIABILITY_Trypanblue.json`

to ISA-JSONv1

```
java -jar enmconvertor.jar -i "INVITRO_VIABILITY_Trypanblue_TEST.xlsx" -j "INVITRO_VIABILITY_Trypanblue.json" -I xlsx -O isa -o "INVITRO_VIABILITY_Trypan blue_TEST.isa.json"
```

Result file (zipped) `INVITRO_VIABILITY_Trypanblue_TEST.isa.json.zip`

to eNanoMapper **RDF** (N3 syntax)

```
java -jar enmconvertor.jar -i "INVITRO_VIABILITY_Trypanblue_TEST.xlsx" -j "INVITRO_VIABILITY_Trypanblue.json" -I xlsx -O isa -o "INVITRO_VIABILITY_Trypanblue_TEST.enm.n3"
```

Result file (zipped) `INVITRO_VIABILITY_Trypanblue_TEST.enm.n3.zip`

[http://ambit.sourceforge.net/enanomapper/
templates/convertor.html](http://ambit.sourceforge.net/enanomapper/templates/convertor.html)

```
{
  "TEMPLATE_INFO": {
    "NAME": "ProteinCorona",
    "VERSION": "TEM",
    "TYPE": 1
  },
  "DATA_ACCESS": {
    "ITERATION": "ROW_SINGLE",
    "SHEET_INDEX": 1,
    "START_ROW": 3,
    "START_HEADER_ROW": 2,
    "END_HEADER_ROW": 2,
    "ALLOW_EMPTY": true,
    "RECOGNITION": "BY_INDEX"
  },
  "SUBSTANCE_RECORD": {
    "SUBSTANCE_UUID": {
      "COLUMN_INDEX": 1
    },
    "PUBLIC_NAME": {
      "COLUMN_INDEX": 1
    },
    "OWNER_NAME": {
      "COLUMN_INDEX": 4
    }
  }
}
```


Conclusion

- Spreadsheets are common
- Spreadsheets are not easy
- We have tools to work with spreadsheets
- Spreadsheets need ontology annotation to be FAIR
- Annotation can be inside the file (e.g. ISA-Tab)
- Annotation can be outside the file (e.g. nmdataparser)