

Proudly Operated by Battelle Since 1965

Capturing and preserving scientific metadata with ISA-TAB

CHASE DOWLING¹, SUSANNA-ASSUNTA SANSONE², NATHAN BAKER¹

¹Pacific Northwest National Laboratory, ²University of Oxford Computer-Aided Drug Design Gordon Research Conference, July 2013

Take-home messages



- Data preservation is important
- Meta-data is a key ingredient to long-term reuse of data
- Most published data is difficult to obtain

An open format exists for data preservation and sharing

Published guidance on model development your and validation: The OECD Principles

To facilitate the consideration of a QSAR model for regulatory purposes, it should be associated with the following information:

- > a defined endpoint
- > an unambiguous algorithm;
- > a defined domain of applicability
- appropriate measures of goodnessof-fit, robustness and predictivity
- > a mechanistic interpretation, if possible;
- >Should be added: data used for modeling should be carefully curated

Slide from Alex Tropsha

Guidelines and associated software tools for reporting, storing, and sharing detailed information considered to be important to include with published data sets on bioactive entities:

NATURE REVIEWS | DRUG DISCOVERY

VOLUME 10 | SEPTEMBER 2011 | 661

OPINION

PERSPECTIVES

Minimum information about a bioactive entity (MIABE)

Sandra Orchard, Bissan Al-Lazikani, Steve Bryant, Dominic Clark, Elizabeth Calder, Ian Dix, Ola Engkvist, Mark Forster, Anna Gaulton, Michael Gilson, Robert Glen, Martin Grigorov, Kim Hammond-Kosack, Lee Harland, Andrew Hopkins, Christopher Larminie, Nick Lynch, Romeena K. Mann, Peter Murray-Rust, Elena Lo Piparo, Christopher Southan, Christoph Steinbeck, David Wishart, Henning Hermjakob, John Overington and Janet Thornton

- A Molecule properties ← (names, structure, InChi, salt, prodrug, ...)
- B Molecule production ← (chemical synthesis, purity, characterization, ...)
- C Physicochemical properties← (molecular weight, water solubility, hydrophobicity, ...)
- □ In vitro cell-free assays ← (primary target, assay details and parameters, delivery systems, secondary gene targets, ...)
- **E** Cellular assays ← (cell type, conditions, assay type, ...)
- F Whole-organism studies ← (animal/plant studies, disease model, toxicology, DDI, ...)
- G Pharmacokinetic studies ← (absorption, dosing route, half-life, Vmax, metabolism, excretion, ...)
 Slide from Alex Tropsha

What is metadata?



Proudly Operated by Battelle Since 1965

- Like Wikipedia pages about Wikipedia
- "Data about data" somewhat ambiguous
 - Structural
 - Descriptive
 - (according to Wikipedia)



Why do we care about...?



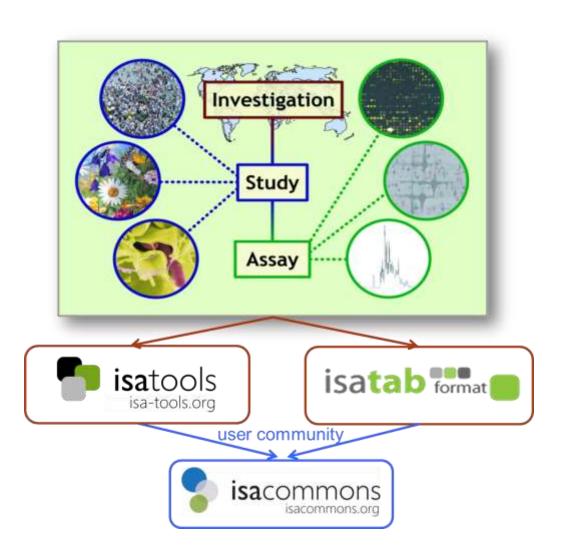
...metadata?

- Your data is not interpretable without it
- Metadata can help to describe unforeseen patterns in data
- It provides features on which to train algorithms
- It improves the speed and accuracy of information retrieval
- ...data preservation?
 - Reproducibility of experiments and calculations
 - Long-term archival of data
 - Increased data accessibility beyond standard journal formats

ISA-TAB: a community-developed data archival format



Proudly Operated by Baffelle Since 1965



ISA-TAB history: community involvement and uptake

1st ISA-Tab workshop 2nd ISA-Tab workshop

3rd ISA-Tab workshop

User workshops/visits - start____1st public instance: _____ Other tools implement ISA-

Tab

Harvard Stem Cell Growing number of Discovery Engine systems starts to adopt ISA framework

Proudly Operated by Battelle Since 1965

Pacific Northwest

Core developments

Straw man ISA-Tab spec

Final ISA-Tab spec

ISA software v1

Database instance at EBI

Conversions to Links to Pride-XML/SRA-XML/-----analysis tools -----> MAGE-Tab starts

RDF/OWL format starts---->

2009

2010

2011

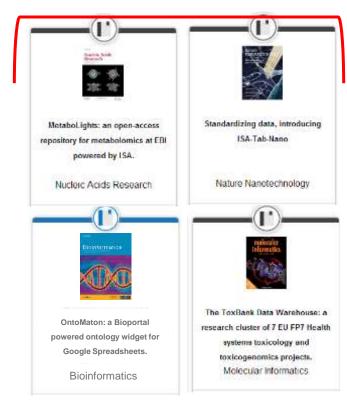
2012

2013

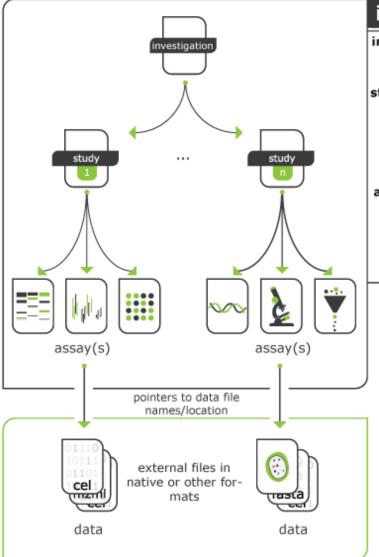
Publications







Proudly Operated by Battelle Since 1965



isa**tab*****

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)

- General-purpose, configurable spreadsheet format, designed to support:
 - use of several omics standards checklists, terminologies
 - reference to CDISC SDTM file(s), and
 - conversions to (a growing number of) other metadata formats,
 - used by public repositories

ISA-TAB community



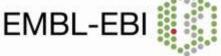


Proudly Operated by Battelle Since 1965

- A grass-root collaborative that works to facilitate collection, curation, and sharing of experiments using a common, structured representation of the experiments that
 - transcends individual biological and technological domains and
 - can be 'configured' to implement several community standards



















Janssen Research & Development, LLC





















Systems Biology of Microorganisms







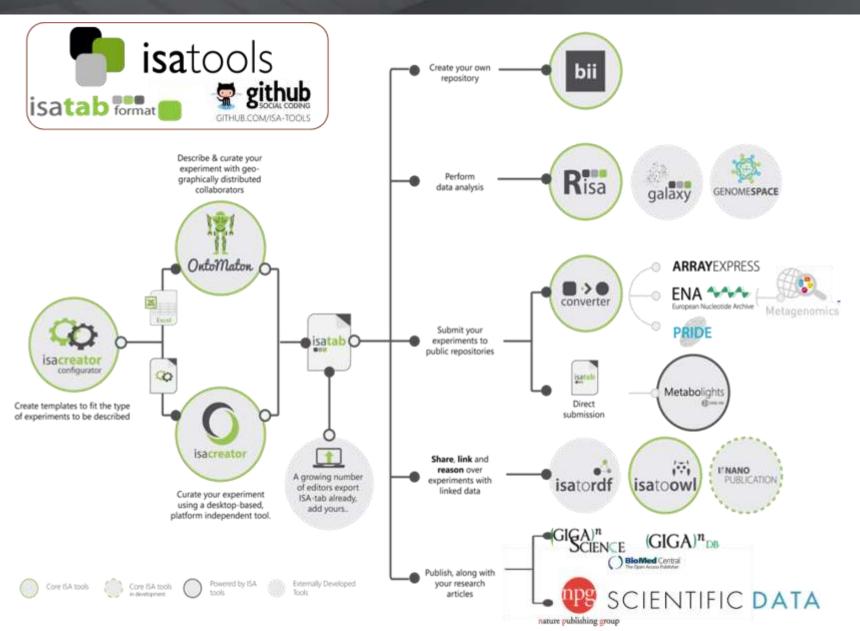




ISA-TAB tools and resources



Proudly Operated by Baffelle Since 1965





Cite the ISA Commons! Towards interoperable bioscience data.

Nature Genetics

2012

Nucleic Acids Research



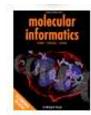


2013

MetaboLights: an open-access repository for metabolomics at EBI powered by ISA.

Nucleic Acids Research





2013

The ToxBank Data Warehouse: a research cluster of 7 EU FP7 Health systems toxicology and toxicogenomics projects.

Molecular Informatics



2013



Nucleic Acids Research

The ISA metadata tracking

framework in action in the Harvard's

Stem Cell Discovery Engine





SCOF STEM CELL DISCOVERY ENGINE



Standardizing data, introducing ISA-Tab-Nano

Nature Nanotechnology



Nanotechnology Informatics Working Group



ISA-TAB applications



- Nanotechnology characterization data
 - National Cancer Informatics Program Nano WG
- Protein titration data
 - pK_a Cooperative

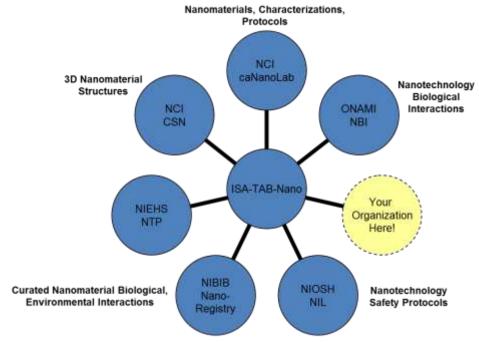
Application to Nanotechnology data



Collaborative work with the NCI Nanotechnology Working Group

- Particularly Stacey Harper, Sharon Gaheen, Dennis Thomas, Mervi Heiskanen, Juli Klemm
- Goal: Develop a specification to facilitate the import/export of data on nanomaterials and their characterizations to/from nanotechnology resources





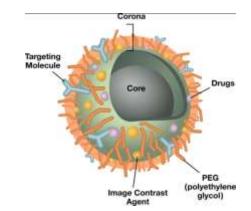
Data sharing challenges for nanotechnology

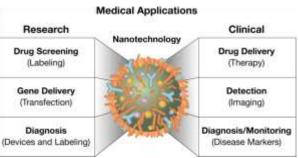


Proudly Operated by Battelle Since 1965

Combinatorial complexity in nanoparticle formulation

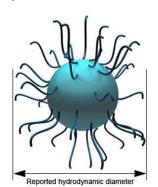
- Diversity of test systems
 - Ecosystem vs. organism vs. cell vs. test tube
 - Species, cell line
 - Age, gender, weight
- Diversity of measurements and assays
 - Physical and chemical: size, potential, surface chemistry, shape, aggregation, ...
 - Biological: toxicity, recognition and association, uptake, delivery, ...
 - Exposure: dose and concentration, timing, duration, ...
- Diversity of data resources with lack of common standard for data exchange





McNeil SE. *J Leukoc Biol*, 2005. **78**(3): p. 585-94.

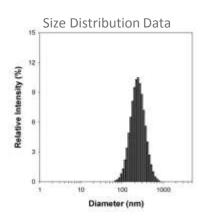
doi:10.1189/jlb.0205074

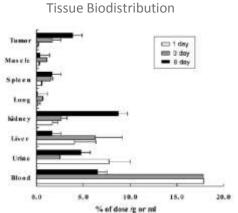


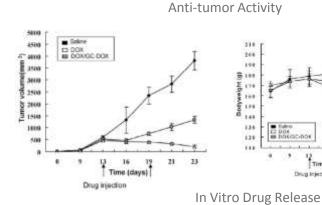
Diversity in assays and data

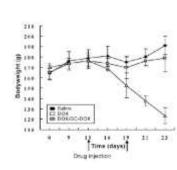


Proudly Operated by Baffelle Since 1965









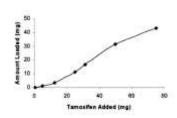
Zeta Potential

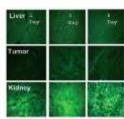
Zeta potential values of control and tamoxifen loaded nano-

hauses	
Nanoparticle formulations	Zeta potential (mV)
Control nanoparticles	6.7±1.29
Turning for loaded nanoparticles	25.4 + 1.4

⁹ Zeta potentials of the nanoparticle suspension in deionized distilled water were measured using the Brookhaven's Zeta PALS instrument.

Drug Loading Data



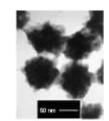


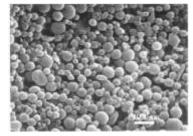
EPC 601 managements for 1 days after 61 approximations as more finally and is a dear of 15 mg/kg. These non-medicine managed by flammacous minimages.

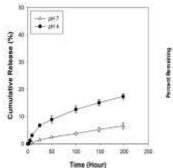
Preparation

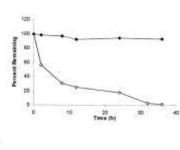
Chemical Composition of Nanoparticle Formulation

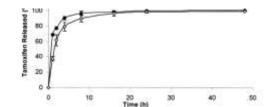
Surface Morphology Data











Source: Chawla JS et al, Int J Pharm, 249, 127-38 (2002), Son YJ et al, J Control Release, 91, 135-145 (2003)

Mean ± S.D. (n = 8).

ISA-TAB-Nano structure

Proudly Operated by Battelle Since 1965

1. Describe the Investigation and Studies

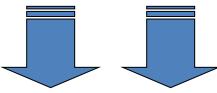
2. Identify Study Samples

3. Record Assay Conditions and Measurements





Investigation File









Material File(s)





Assay File(s)

ISA-TAB-Nano Investigation File



Describes:

- Primary investigation
- Associated materials, studies, assays, and protocols
- Descriptive information about the study includes:
 - Design descriptors and factors
 - Publications
 - Assays and protocols
 - Contacts
- Vertical-based spreadsheet format with columns representing multiple values

ISA-TAB-Nano Investigation File

ONTOLOGY SOURCE REFERENCE

Investigation Person Affiliation

Investigation Person Roles Term Accession Number Investigation Person Roles Term Source REF

Investigation Person Roles



Proudly Operated by Baffelle Since 1965

(ON TOLOGY SOUNCE REPERENCE		
	Term Source Name	MO	NPO
Ontology	Term Source File	http://purl.bioontology.org/ontology/MO	http://purl.bioontology.org/ontology/npo
References	Term Source Version	v. 1.3.1.1	v. 2011-02-12
reservinces	Term Source Description	MGED Ontology	NanoParticle Ontology
l	INVESTIGATION		
	Investigation Identifier	NCL200612A	
(Investigation Title	Dendrimer-Based MRI Contrast Agents	
I		The goal of this investigation is to characterize	1
		PAMAM dendrimer with an associated	
		gadolinium chelate MRI contrast agent.	
Investigation {	Investigation Description		
- 1	Investigation Submission Date	2002-11-30	
Description	Investigation Public Release Date	2002-11-30	
	Investigation Disease		
	Investigation Disease Term Accession Number		
(Investigation Disease Term Source REF		
;	Investigation Outcome		
(INVESTIGATION PUBLICATIONS		
	Investigation PubMed ID	18095846	
	Investigation Publication DOI	10.2217/17435889.2.6.789	
Investigation	Investigation Publication Author List	10.2217/17435889.2.6.789	
Publications		Characterization of nanoparticles for	
Publications)	Investigation Publication Title	therapeutics	
	Investigation Publication Status	published	
	Investigation Publication Status Term Accession Number		
(Investigation Publication Status Term Source REF		
`	INVESTIGATION CONTACTS		
(Investigation Person Last Name	Doe	
	Investigation Person First Name	John	
	Investigation Person Mid Initials	E	
	Investigation Person Email	doej@mail.nih.gov	
Investigation	Investigation Person Phone	1231231234	
- ,	Investigation Person Fax		
Contacts 1	N. C.		
	Investigation Person Address	Laboratory Street, City, State 111111	

July 27, 2013

MO

Doe Laboratories

investigator

ISA-TAB-Nano Study File



- Provides mapping between the study samples, materials, and processing events
- Samples can be:
 - Biological materials
 - Nanomaterials
 - Small molecules
- For physical-chemical characterizations of nanomaterials, the sample is the nanomaterial
- For in vitro and in vivo characterizations, the sample is the biological specimen (cell line, animal, etc.)
- Horizontal spreadsheet describing the biological materials and association with the nanomaterials described in the Material file

ISA-TAB-Nano Study File



Proudly Operated by Battelle Since 1965

В	İΟ	log	ıical	Sam	ple	lden	tifie

Biological Sample Characteristics

Protocol

		-				
Source Name	Material Type	Characteristics[cell type{NCIt:C12508}]	Characteristics[ATCC#{NCIt:C15661}]Provider	Protocol REF	Performer
LLC-PK1	biospecimen	porcine proximal tubule cells	CL-101	Doe Technologies	cell preparation in four 96-well plates	Jane Doe
LLC-PK1	biospecimen	porcine proximal tubule cells	CL-101	Doe Technologies	cell preparation in four 96-well plates	Jane Doe
LLC-PK1	biospecimen	porcine proximal tubule cells	CL-101	Doe Technologies	cell preparation in four 96-well plates	Jane Doe
LLC-PK1	biospecimen	porcine proximal tubule cells	CL-101	Doe Technologies	cell preparation in four 96-well plates	Jane Doe
LLC-PK1	biospecimen	porcine proximal tubule cells	CL-101	Doe Technologies	cell preparation in four 96-well plates	Jane Doe
LLC-PK1	biospecimen	porcine proximal tubule cells	CL-101	Doe Technologies	cell preparation in four 96-well plates	Jane Doe
LLC-PK1	biospecimen	porcine proximal tubule cells	CL-101	Doe Technologies	cell preparation in four 96-well plates	Jane Doe
LLC-PK1	biospecimen	porcine proximal tubule cells	CL-101	Doe Technologies	cell preparation in four 96-well plates	Jane Doe
LLC-PK1	biospecimen	porcine proximal tubule cells	CL-101	Doe Technologies	cell preparation in four 96-well plates	Jane Doe

Biological Sample

Study Factors

		Factor Value[particle				Factor Value[time of	
Sample Name	Factor Value[nanoparticle sample]	concentration]	Unit	Term Accession Number	Term Source Ref	exposure]	Unit
LLC-PK1-6h-NCL22-1	NCL-22	0.004	mg/mL	UO_0000176	UO	6	hour
LLC-PK1-6h-NCL22-2	NCL-22	0.008	mg/mL	UO_0000176	UO	6	hour
LLC-PK1-6h-NCL22-3	NCL-22	0.016	mg/mL	UO_0000176	UO	6	hour
LLC-PK1-6h-NCL22-4	NCL-22	0.032	mg/mL	UO_0000176	UO	6	hour
LLC-PK1-6h-NCL22-5	NCL-22	0.064	mg/mL	UO_0000176	UO	6	hour
LLC-PK1-6h-NCL22-6	NCL-22	0.128	mg/mL	UO_0000176	UO	6	hour
LLC-PK1-6h-NCL22-7	NCL-22	0.256	mg/mL	UO_0000176	UO	6	hour
LLC-PK1-6h-NCL22-8	NCL-22	0.512	mg/mL	UO_0000176	UO	6	hour
LLC-PK1-6h-NCL22-9	NCL-23	1	mg/mL	UO_0000176	UO	6	hour

ISA-TAB-Nano Material File



- Primary file for describing:
 - Nanomaterial composition and formulation
 - Physical properties
 - Structure
- Allows for:
 - Comparison of nanomaterials across nanotechnology resources
 - Association with optional files; e.g., a Structure file for representing the 3D structure of the nanomaterial
- Horizontal spreadsheet describing the nanomaterial sample, associated components, material characteristics, and material linkages

ISA-TAB-Nano Material File



Proudly Operated by Battelle Since 1965

Material	Identifiers	Mate	Intended Application		
				Material Design	Material Intended
Material Source Name	Material Name	Material Description	Material Synthesis	Rationale	Application
NCL-22	g45_coona_dendrimer	G4.5 COONa terminated PAMAM dendrimer			delivery of image contrast agent
NCL-23	g45_coona_dendrimer_ma gnevist_complex	G4.5 COONa terminated PAMAM dendrimer-Magnevist® complex			MRI
NCL-24	magnevist	gadolinium based image contrast agent			MRI contrast agent

	Material Type		Che	Characteristics		
			Material Chemical	Term Accession		Characteristics[dendrimer
Material Type	Term Accession Number	Term Source REF	Name	Number	Term Source REF	branch]
dendrimer; conjugated	NDO 735, NDO 1936	NPO;NPO				1 4
component	NPO_735; NPO_1826	NPO;NPO				1-4
nanoparticle sample	NPO_1404	NPO				
mall molecule; imaging						
payload agent; conjugated	NCIt_C48809;		gadopentetate			
omponent	NPO_1534; NPO_1826	NCIt;NPO;NPO	dimeglumine	31797	ChEBI	

ISA-TAB-Nano Material File



Proudly Operated by Battelle Since 1965

Characteristics

					•
Characteristics[dendrimer generation]	Characteristics[molecular weight]	Unit	Term Accession Number	Term Source REF	Characteristics[molecular formula]
4.5	26.28	kDa	UO_0000222	UO	
					[Gd+3] .CNC[C@H] (O) [C@@H] (O) [C@H] (O) [C@H] (O) CO.CNC[C@H] (O) [C@@H] (O) [C@H] (O) [C@H] (O) CO.OC(=O) CN(CCN(CCN(CC(O) =O) CC([O-]) =O) CC([O-]) =O) CC([O-]) =O

Material Linkage

Material File

					•					
			Term			Material			Material	Material Data
Material	Material Linkage	Material Linkage	Accession	Term Source	Material Data	Data File	Term Accession	Term Source	Data File	File
Constituent	Name	Туре	Number	REF	File	Туре	Number	REF	Version	Description
g45_coona_dend										
rimer; magnevist		covalent linkage	NPO_563	NPO	magnevist.jpg	image	NCIt_C48179	NCIt		

ISA-TAB-Nano Assay File



Proudly Operated by Battelle Since 1965

- Describes the protocol parameters and factors, including:
 - Temperature
 - Media/solvent
 - Concentration
- Provides references or links to assay results, including:
 - Measurements
 - Instrumentation
 - Derived data files
- Templates available for the "top Nano WG assays"
 - Size by DLS (Physico-Chemical)
 - Zeta Potential (Physico-Chemical)
 - Hemolysis (In Vitro)
 - Hepatocarcinoma Cytoxicity (MTT and LDH) (In Vitro)
 - Caspase 3 Apoptosis (In Vitro)
 - Toxicity (ADME, Single/Repeat Dose) (In Vivo)
 - Your assay here!

ISA-TAB-Nano Assay File



Proudly Operated by Battelle Since 1965

Sample Name	S
-------------	---

Protocol Parameters

Assay Factors

Sample Name	Protocol REF	Performer	Assay Name	Factor Value[temperature]	Term Accession Number	Term Source Ref	Unit	Term Accession Number	Term Source Ref	Factor Value[solvent medium]
NCL-20-1	Measuring the size of nanoparticles in aqueous media using batch-mode dynamic light scattering	John Doe	size by DLS assay	25	PATO_0000146	PATO .	celsius	UO_0000027	uo	saline
NCL-20-1	Measuring the size of nanoparticles in aqueous media using batch-mode dynamic light scattering	John Doe	size by DLS assay	25	PATO_0000146	PATO	celsius	UO_0000027	uo	PBS
NCL-22-1	Measuring the size of nanoparticles in aqueous media using batch-mode dynamic light scattering	John Doe	size by DLS assay	25	PATO_0000146	PATO	celsius	UO_0000027	uo	saline
NCL-22-1	Measuring the size of nanoparticles in aqueous media using batch-mode dynamic light scattering	John Doe	size by DLS assay	25	PATO_0000146	SPATO	celsius	UO_0000027	uo	PBS
NCL-22-1	Measuring the size of nanoparticles in aqueous media using batch-mode dynamic light scattering		size by DLS assay	37	PATO_0000146	PATO .	celsius	UO_0000027	uo	PBS

Assay Measurements

Assay Files

									•
Unit	Term Accession Number	Term Source Ref	Measurement Value[peak size]	Unit	Term Accession Number	Term Source Ref	Measurement Value[pdi]	Derived Data File	
								NCL-Dendrimer-	
nm	UO_000018	UO	4.4	nm	UO_000018	UO	0.122	Based_MRI_Contrast_Agent.pdf	
								NCL-Dendrimer-	
nm	UO_000018	UO	6.2	nm	UO_000018	UO	0.211	Based_MRI_Contrast_Agent.pdf	
								NCL-Dendrimer-	
nm	UO_000018	UO	6	nm	UO_000018	UO	0.2	Based_MRI_Contrast_Agent.pdf	
								NCL-Dendrimer-	
nm	UO_000018	UO	5.2	nm	UO_000018	UO	0.214	Based_MRI_Contrast_Agent.pdf	
								NCL-Dendrimer-	
nm	UO_000018	UO	5.1	nm	UO_000018	UO	0.282	Based_MRI_Contrast_Agent.pdf	
	nm nm nm	nm UO_0000018 nm UO_0000018 nm UO_0000018 nm UO_0000018	Number Ref nm UO_0000018 UO nm UO_0000018 UO nm UO_0000018 UO nm UO_0000018 UO	Number Ref Value[peak size] nm UO_0000018 UO 4.4 nm UO_0000018 UO 6.2 nm UO_0000018 UO 6 nm UO_0000018 UO 5.2	Number Ref Value[peak size] nm UO_0000018 UO 4.4 nm nm UO_0000018 UO 6.2 nm nm UO_0000018 UO 6 nm nm UO_0000018 UO 5.2 nm	Number Ref Value[peak size] Number nm UO_0000018 UO 4.4 nm UO_0000018 nm UO_0000018 UO 6.2 nm UO_0000018 nm UO_0000018 UO 6 nm UO_0000018 nm UO_0000018 UO 5.2 nm UO_0000018	Number Ref Value[peak size] Number Ref nm UO_0000018 UO 4.4 nm UO_0000018 UO nm UO_0000018 UO 6.2 nm UO_0000018 UO nm UO_0000018 UO 6 nm UO_0000018 UO nm UO_0000018 UO 5.2 nm UO_0000018 UO	Number Ref Value[peak size] Number Ref Value[pdi] nm UO_0000018 UO 4.4 nm UO_0000018 UO 0.122 nm UO_0000018 UO 6.2 nm UO_0000018 UO 0.211 nm UO_0000018 UO 6 nm UO_0000018 UO 0.2 nm UO_0000018 UO 5.2 nm UO_0000018 UO 0.214	Number Ref Value[peak size] Number Ref Value[pdi] nm UO_0000018 UO 4.4 nm UO_0000018 UO 0.122 Based_MRI_Contrast_Agent.pdf nm UO_0000018 UO 6.2 nm UO_0000018 UO 0.211 Based_MRI_Contrast_Agent.pdf nm UO_0000018 UO 6 nm UO_0000018 UO 0.2 Based_MRI_Contrast_Agent.pdf nm UO_0000018 UO 5.2 nm UO_0000018 UO 0.214 Based_MRI_Contrast_Agent.pdf NCL-Dendrimer- NCL-Dendrimer- NCL-Dendrimer- NCL-Dendrimer- NCL-Dendrimer-

ISA-TAB-Nano References and Team



Proudly Operated by Battelle Since 1965

- ► ISA-TAB-Nano Project Site: <u>https://wiki.nci.nih.gov/display/ICR/ISA-TAB-Nano</u>
- ASTM standard: http://www.astm.org/Standards/ E2909.htm
- ISA-TAB: http://isa-tools.org
- caBIG ICR Nano WG Data Standards Document: https://wiki.nci.nih.gov/display/learthgov/d
- NanoParticle Ontology (NPO): http://www.nano-ontology.org

ISA-TAB-Nano Project Team

Nathan Baker, PNNL **Dennis Thomas. PNNL** Amy Bednar, ERDC Elaine Freund, 3rd Millennium **Marty Fritts, NCL** Sharon Gaheen, SAIC Sue Pan, SAIC Liz Hahn-Dantona, Lockheed Martin Stacey Harper, Oregon State University Mark Hoover, NIOSH Fred Klaessig, Pennsylvania Bio Nano Systems Juli Klemm, NCI CBIIT Mervi Heiskanen, NCI CBIIT **David Paik, Stanford University Grace Stafford, The Jackson Laboratory** Todd Stokes, Georgia Tech

Application to protein titration data

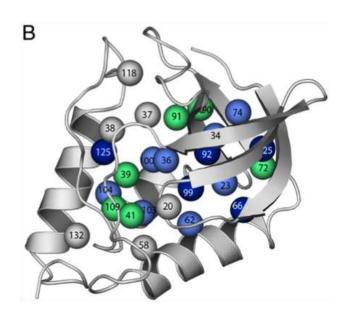


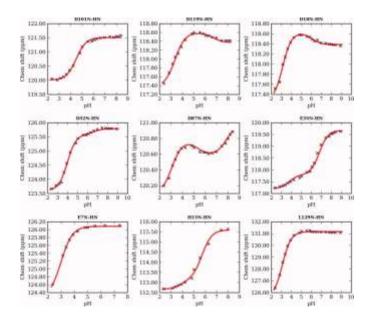
Proudly Operated by Baffelle Since 1965

Collaborative work with the pK_a Cooperative

Particularly Chase Dowling, Jens Nielsen, Bertrand Garcia-Moreno, Marilyn Gunner, Anthony Nicholls

Goal: Preserve experimental measurements and computational predictions of protein pK_a data for benchmarking and improvement of biomolecular electrostatics models

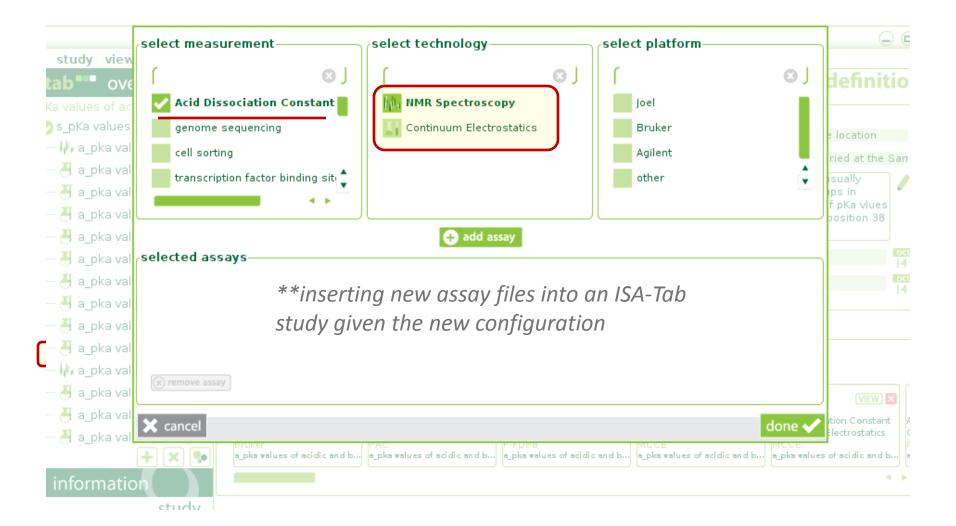




Configuring ISA-Tab for pK_a data

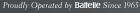


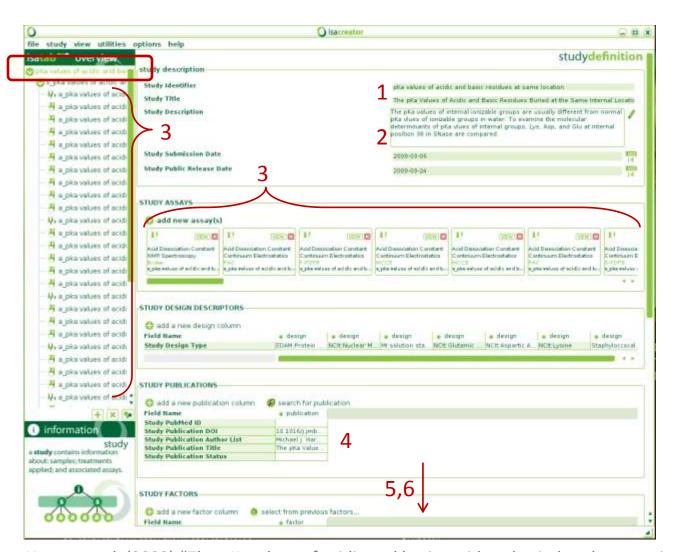
Addition of configurations for NMR Spectroscopy and Continuum Electrostatics for structural chemistry assays – assignment of required values, platform/software used, etc.



Investigation/Study/Assay







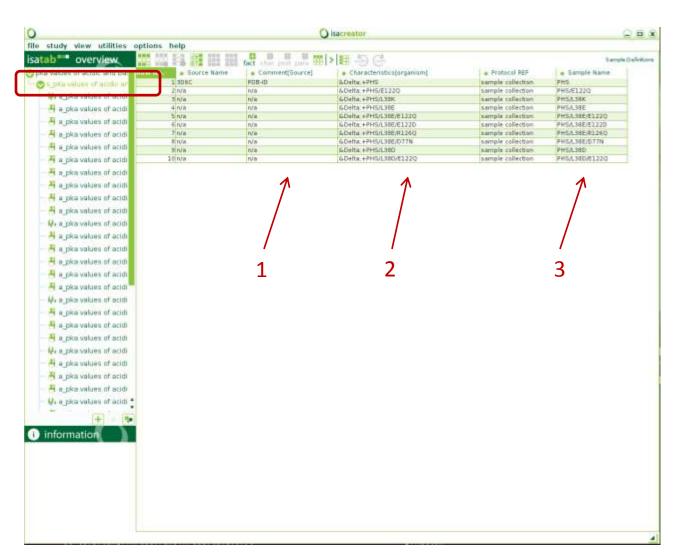
Investigation File: Current level circled

- 1. Study (journal article) Title
- 2. Abstract
- 3. Assay Files
- 4. Digital Object Information
- 5. Study Information
- 6. Author/Contact Information

Harms, et al. (2009) "The pK_a values of acidic and basic residues buried at the same internal location are governed by different factors" (Garcia-Moreno lab)

Investigation/Study/Assay





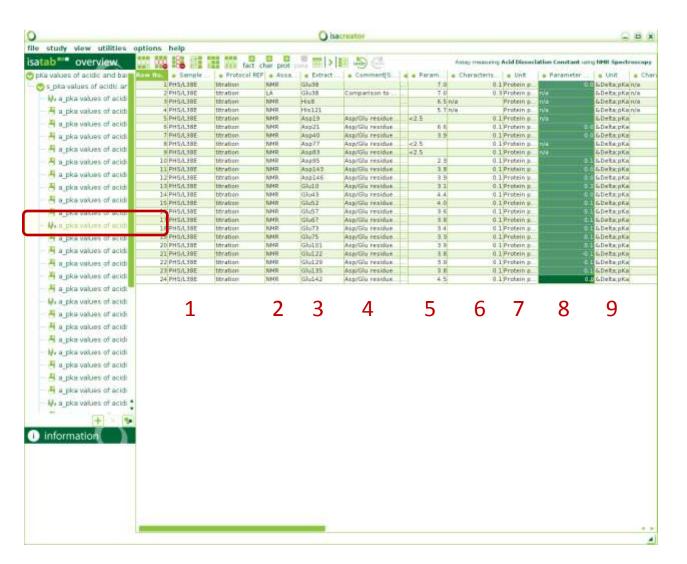
Study File: Current level circled

- 1. Protein Source
- 2. Protein Mutants
- 3. Protein Mutant Sample Reference Names

This example is looking at the Δ +PHS variant of Staphylococcal Nuclease along with various residue replacements. The PDB-ID is given for the base variant.

Investigation/Study/Assay





Example Assay File: Current level circled

- Sample Reference Name (L38E replacement)
- Assay type NMR Spectroscopy
- 3. Buried residue being examined
- 4. Significant comment included from journal article data table for specific values
- 5. Measured pKa value
- 6. Standard deviation
- 7. Unit
- Calculated change in pKa from other reference value
- 9. Unit

List of currently converted datasets



- Data available on the pKa Coop website: http://pkacoop.org/
 - Harms, et al (2009) "The pKa values of acidic and basic residues buried at the same internal location in a protein are governed by different factors"
 - Castaneda, et al (2009) "Molecular determinants of the pKa values of Asp and Glu residues in Staphylococcal nuclease"
 - Harms, et al (2008) "A buried lysine that titrates with a normal pKa: Role of conformational flexibility at the protein—water interface as determinant of pKa values"
 - Fitch, et al (2002) "Experimental pKa values of buried residues: Analysis with continuum methods and role of water penetration"
 - Perez-Canadillas, et al (1998) "Characterization of pKa values and titration shifts in cytotoxic ribonuclease α-sarcin by NMR. Relationship between electrostatic interactions, structure, and catalytic function"
 - Czodrowski (2011) "Blind, one-eyed, or eagle-eyed? pKa calculations during blind predictions with staphylococcal nuclease"
 - Gunner and Zheng (2008) "Analysis of the electrochemistry of hemes with Ems spanning 800 mV"
 - Further articles for curation cited in Song, Mao, and Gunner (2009) "MCCE2: Improving protein pKa calculations with extensive side chain rotamer sampling"

Next steps for protein pK_a data preservation



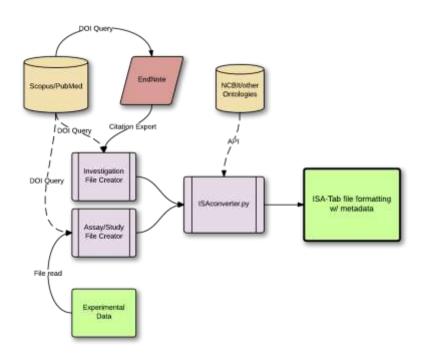
Proudly Operated by Battelle Since 1965

Currently

- Accelerating curation with new tools
- Feeding DOI's of articles measuring protein residue pKa's into ISApy
- Improving flexibility of data file conversion

Future

- Integrating data warehousing and analysis
- Working with publishers on standard data formatting
- Expanding ISApy for all assays in the default configuration



Take-home messages



- Data preservation is important
- Meta-data is a key ingredient to long-term reuse of data
- Most published data is difficult to obtain

An open format exists for data preservation and sharing

Acknowledgments



- Collaborators
 - pKa Cooperative
 - National Cancer Informatics Program Nanotechnology Working Group
 - ISA-TAB Team
- Funding
 - National Cancer Informatics Program Nanotechnology Working Group
 - NIH R01 GM069702, U01 NS073457-01
 - OpenEye Software



Chase Dowling