Data Collection from the Nanotoxicology Literature Using ISA-TAB-Nano

R.L. Marchese Robinson, A.-N. Richarz, A. Cassano, M.T.D. Cronin, R. Rallo

Presentation delivered to the U.S. Nano WG

21st of January 2016

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Overview

- Why does (meta)data standardisation matter?
- Brief introduction to ISA-TAB-Nano
- Summary of NanoPUZZLES work (2013-2015)
- ISA-TAB-Nano challenges
- Outlook
- Conclusions

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Why does (meta)data standardisation matter?

- Differently structured datasets => new code or configuration files
- Lack of standardised terminology => are we talking about the same thing?
  - Ontologies explicitly define concepts and relationships
- Standardisation supports data integration & computational analysis

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

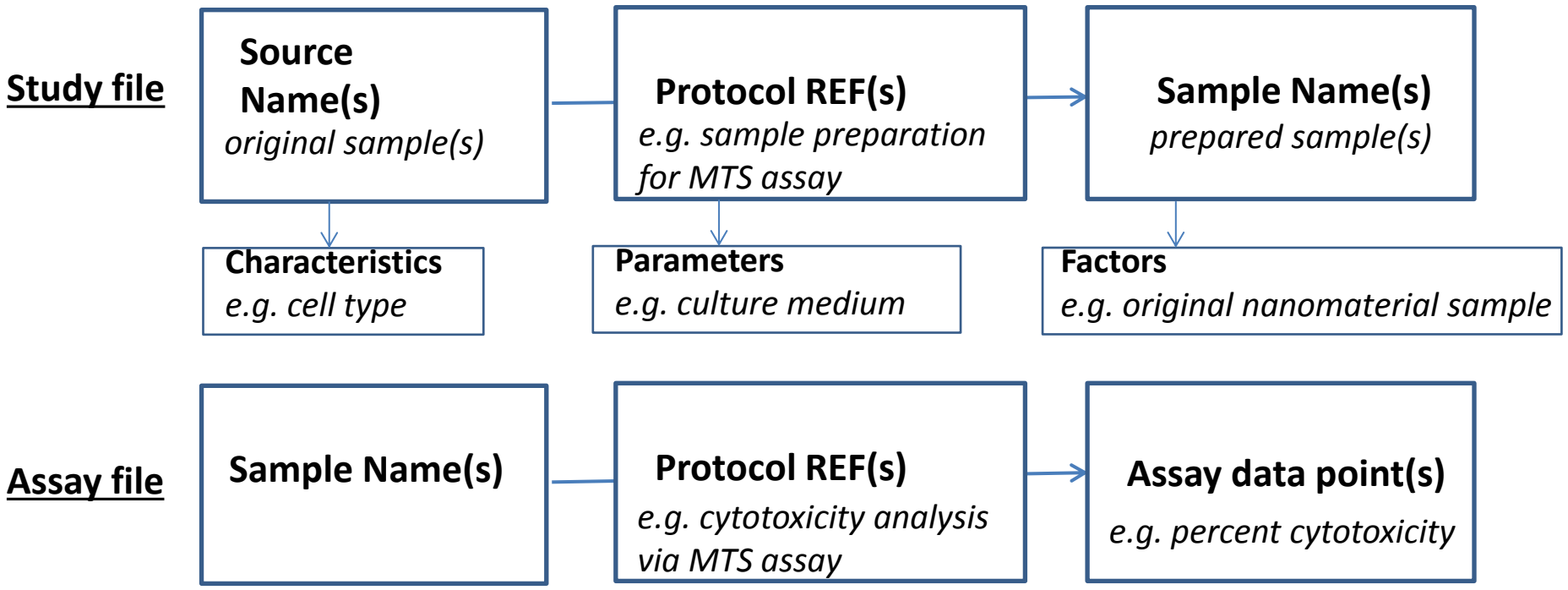
ISA-TAB-Nano: motivation and current status[1,2]

- Proposed nanoscience community data EXCHANGE standard
  - Can record or link to data
  - Metadata standardisation PROMOTED
    - e.g. ontology links supported
- Interconnected "spreadsheet-like" file *types*
  - **I**nvestigation, **S**tudy, **A**ssay, Material – **TAB**ular files
  - Flexible – *not* fully specified fields, but field *types* e.g. factors
  - Business rules promote standardised addition of fields
- Iterative development
  - Adapted from ISA-TAB[3]
  - Original publication in 2013[1]
  - Current version: 1.2[2]

1. Thomas, D. G. et al. *BMC Biotechnol.*, *13*, **2013**, 2.
2. https://wiki.nci.nih.gov/display/ICR/ISA-TAB-Nano
3. http://www.isa-tools.org/format/specification/

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

ISA-TAB-Nano: data model



- Sometimes a judgement call required as to which experimental variables should be described via which field types: characteristics, parameters, factors

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

ISA-TAB-Nano: Material file[1]

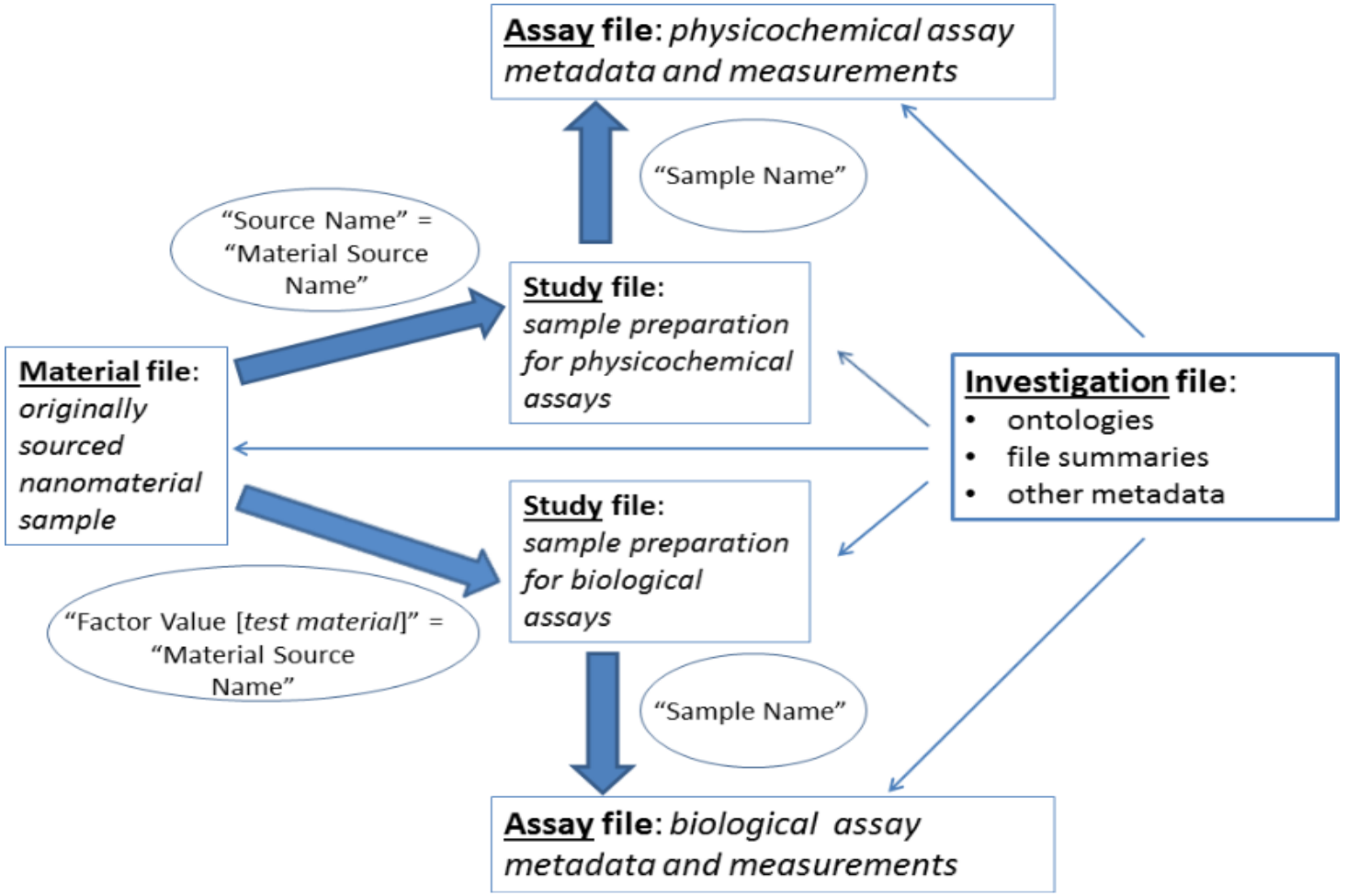
Material Source Name	Material Name	Material Type	Material Chemical Name	Characteristics [nominal size]	Unit
original sample ID	original sample ID	core/ shell NP	coated TiO2	10	nm
original sample ID	part_1	core	TiO2		
original sample ID	part_2	shell	SiO2		

- Some fields are omitted

1. https://wiki.nci.nih.gov/display/ICR/Material

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

ISA-TAB-Nano: linked files[1]

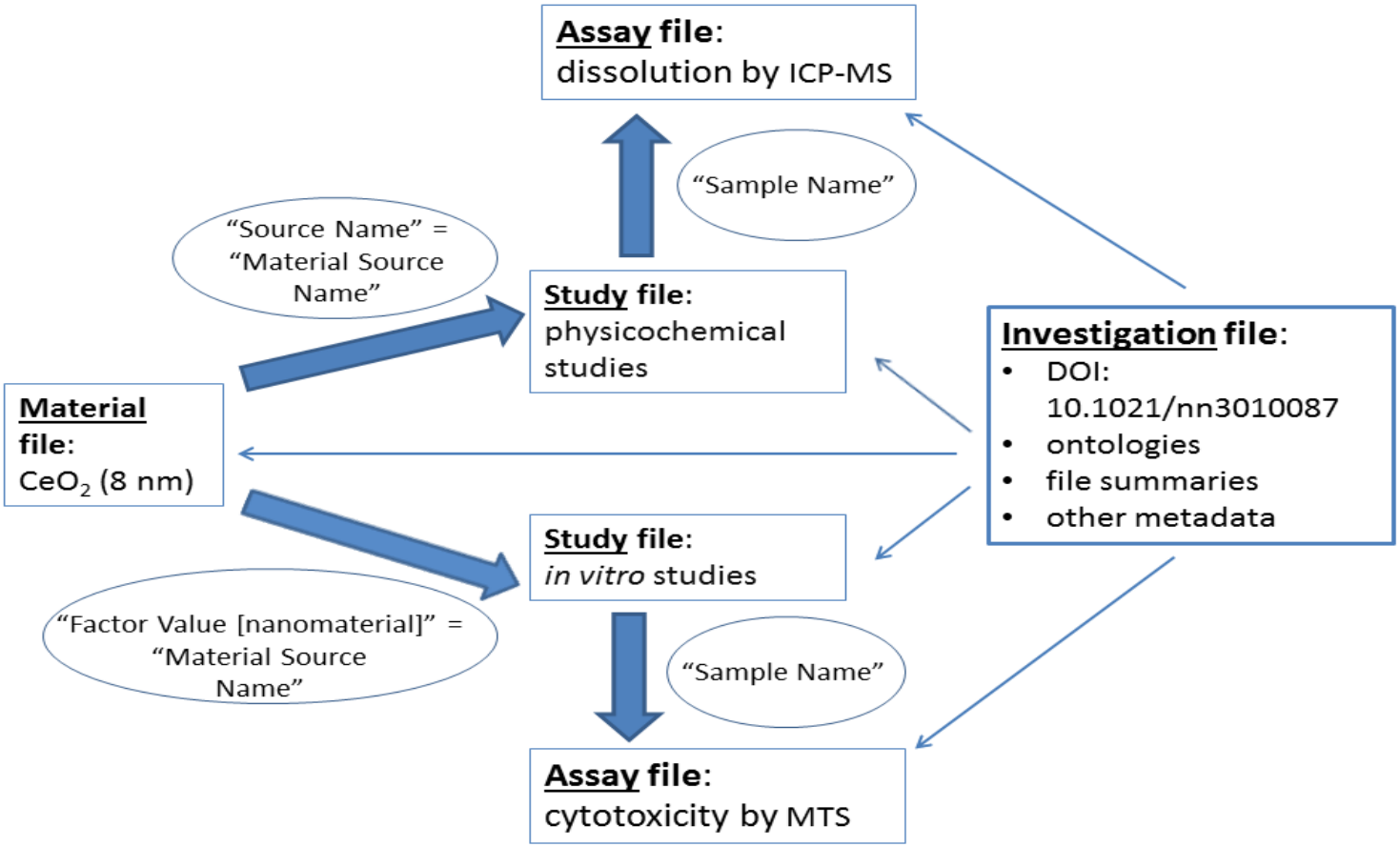


- "Factor Value [*test material*]" *e.g.* "Factor Value [nanomaterial]"

1. Figure 1: Marchese Robinson, R.L. et al. *Beilstein J. Nanotechnol.*, *6*, **2015**, 1978–1999.

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

NanoPUZZLES ISA-TAB-Nano data collection: example



- Subset of NanoPUZZLES dataset derived from Zhang et al.[1]

1. Zhang, H. et al. *ACS Nano*, *6*, **2012**, 4349-4368.

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

NanoPUZZLES ISA-TAB-Nano data collection: summary (1)

- NanoPUZZLES project: 2013 – 2015[1]
- Nanosafety focus
- Biological endpoints: cytotoxicity, genotoxicity, (embryo) zebrafish mortality
- Physicochemical endpoints: zeta potential, size, shape, dissolution, adsorption, surface area, crystal phase
- More than 200 (nominal) nanomaterials
  - e.g. metals, (metal) oxides, carbon nanotubes, fullerene

1. http://www.nanopuzzles.eu

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

NanoPUZZLES ISA-TAB-Nano data collection: summary (2)

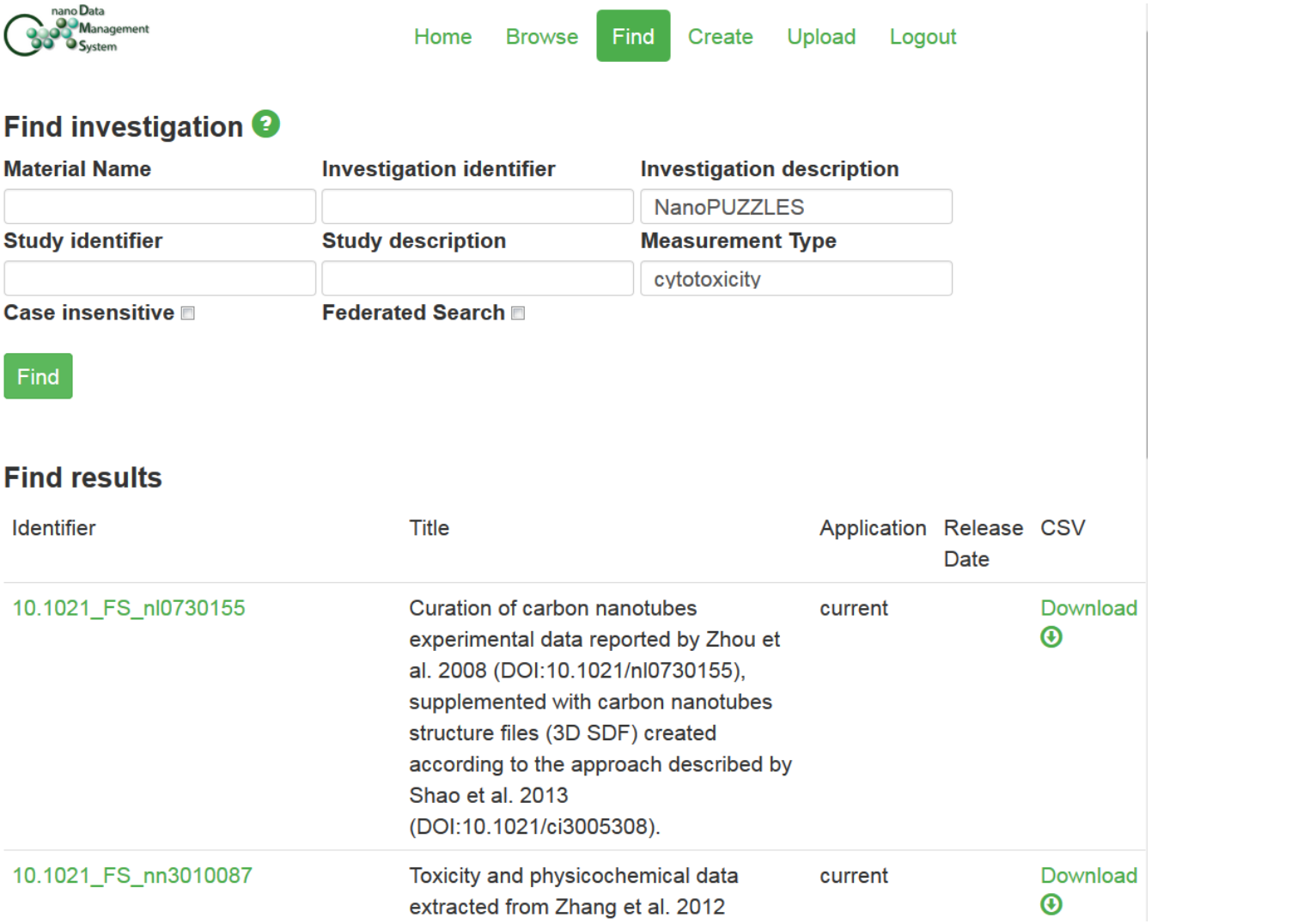
- Publicly released on Zenodo[1] and FigShare[2] (end of December 2015)
- Creative Commons Attribution license
- Cytotoxicity and genotoxicity datasets submitted to nanoDMS database (MODERN project)[3]
  - see next slide

1. http://dx.doi.org/10.5281/zenodo.35493
2. https://figshare.com/search?q=NanoPUZZLES&quick=1
3. http://biocenitc-deq.urv.cat/nanodms

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

NanoPUZZLES ISA-TAB-Nano data collection: summary (3)

MODERN project nanoDMS database integration:



This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6] Grant Agreement nr 309837

NanoPUZZLES ISA-TAB-Nano data collection tools

- Data collection tools: iteratively updated (Excel based) templates,[1] business rules,[2] software[3]
  - software creates text files e.g. for nanoDMS upload[4]
  - close collaboration with MODERN enabled compatibility
- A snapshot of these tools was recently described in detail[5]
- NanoPUZZLES data collection tools partially and imperfectly addressed some challenges[5]

1. http://www.myexperiment.org/files/1356.html
2. Some recent additions beyond [5] documented in dataset README file
3. https://github.com/RichardLMR/xls2txtISA.NANO.archive
4. http://biocenitc-deq.urv.cat/nanodms
5. Marchese Robinson, R.L. et al. *Beilstein J. Nanotechnol.*, *6*, **2015**, 1978–1999.

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Challenges: overview

- Various challenges associated with ISA-TAB-Nano data collection from the nanotoxicology literature were recently discussed
  - e.g. some in Marchese Robinson et al.[1]
- Not all of these are specific to ISA-TAB-Nano
  - e.g. minimum information standards for nanoscience
- Not all of these require changes to ISA-TAB-Nano functionality
- Many require agreement on and explicit documentation of best practice

1. Marchese Robinson, R.L. et al. *Beilstein J. Nanotechnol.*, *6*, **2015**, 1978–1999.

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Challenges: experimentally determined chemical composition

Material Source Name	Material Type	Material Chemical Name	Characteristics [nominal size]	Unit
original sample ID	core/ shell NP	coated TiO2	10	nm

- As of version 1.2, experimentally determined “Characteristics” => Assay file entries[1,2]
- What about experimentally determined components, “Material Chemical Name”, “Material Linkage Type” etc.?
- What about experimentally determined “Characteristics” linked to a specific component e.g. impurity percentage in a shell?

1. https://wiki.nci.nih.gov/display/ICR/Material
2. https://wiki.nci.nih.gov/display/ICR/ISA-TAB-Nano%201.2%20Release%20Notes

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Challenges: mixtures

Study file

Factor Value [exposure medium serum]	Term Accession Number	Term Source REF	Factor Value [exposure medium serum concentration]	Unit
fetal bovine serum; horse serum	http://.....#CCONT_0000048; http://.....#CCONT_0000058	CCONT; CCONT	10;8	percent; percent

- For example, serum = 10% fetal bovine serum, plus 8% horse serum
- NanoPUZZLES approach illustrated – but imperfect[1]
- ISA-TAB developers have recently proposed a different solution (details forthcoming)

1. Marchese Robinson, R.L. et al. *Beilstein J. Nanotechnol.*, *6*, **2015**, 1978–1999.

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Outlook: Nanoinformatics community

- ISA-TAB-Nano specification under review
  - work started towards version 1.3
  - ongoing discussions
  - partly prompted by issues identified within NanoPUZZLES
  - will aim to harmonise with FORTHCOMING revised ISA-TAB specification

- Proposed development of community accepted templates
  - initial suggestion
  - to be led by ISA-TAB-Nano developers

- Resources being developed by other projects
  - e.g. tools for creating and parsing ISA-TAB-Nano files under development within eNanoMapper
  - e.g. KNIME workflows for analysing ISA-TAB-Nano datasets under development within MODERN

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Conclusions

- Best use of ISA-TAB-Nano will be facilitated by
  - clarification and explicit documentation of best practice and resolution of other challenges
  - development of harmonised templates
  - further development of software to facilitate creation and analysis

- The work carried out within NanoPUZZLES provides some useful insights which will support this broader work
  - not a definitive solution

- The NanoPUZZLES datasets will be a useful data resource for the community

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Acknowledgements

– Funding

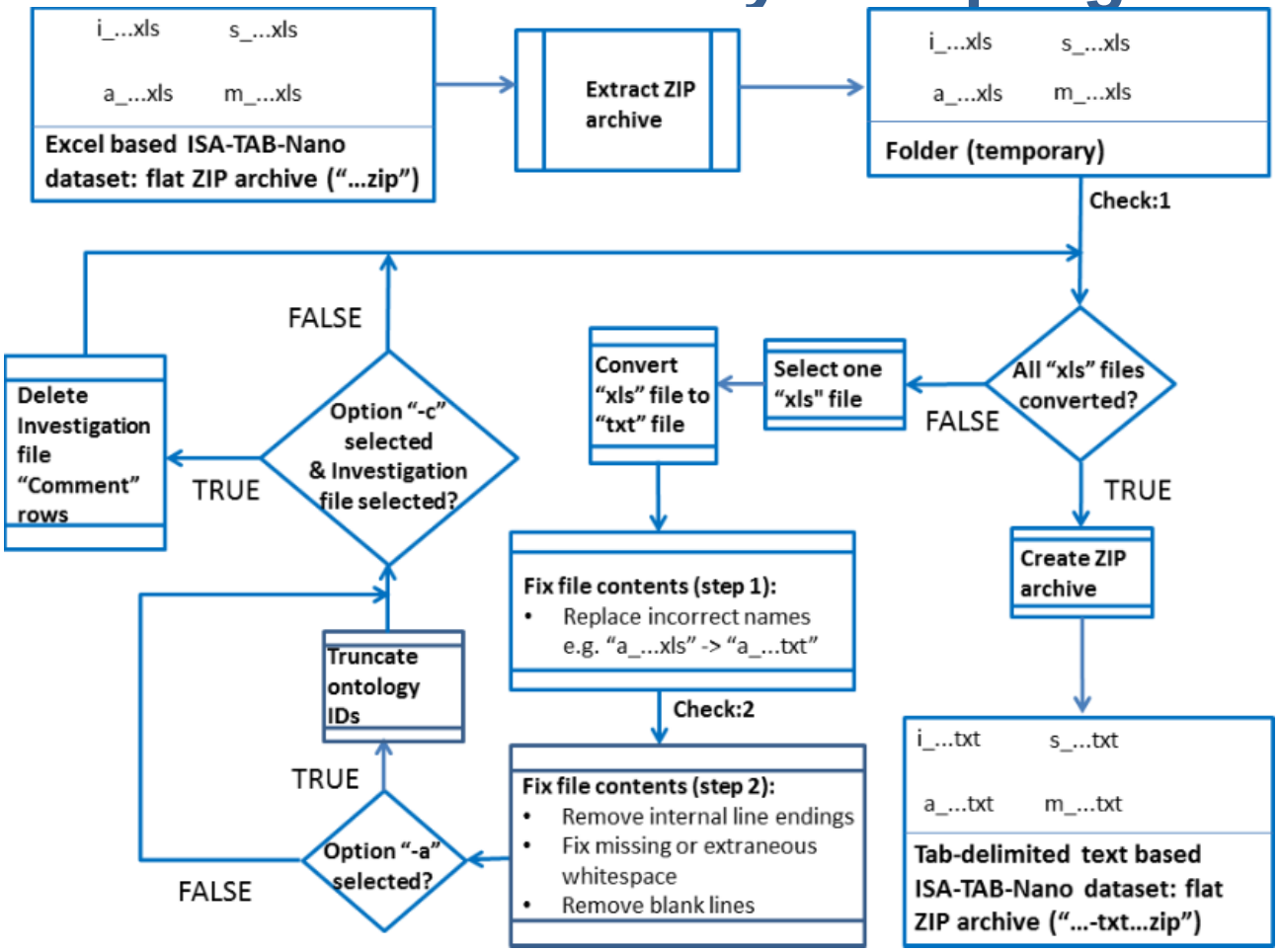
The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/ 2007-2013) under grant agreements no. 309837 (NanoPUZZLES project) and no. 309314 (MODERN project)

– Useful interactions

- NSC Databases Working Group
- US Nano WG
- Sharon Gaheen (Leidos Biomedical Research Inc.)
- Nathan Baker (Pacific Northwest National Laboratory)
- Nina Jeliazkova (IdeaConsult Ltd.)
- Philippe Rocca-Serra (University of Oxford)
- Christoffer Åberg (University of Groningen)
- Neill Liptrott (University of Liverpool)
- Claire Mellor (Liverpool John Moores University)
- Rafi Korenstein (Tel-Aviv University)
- Lang Tran and Peter Ritchie (Institute of Occupational Medicine)
- Roger Pons and Josep Cester (Universitat Rovira i Virgili)

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

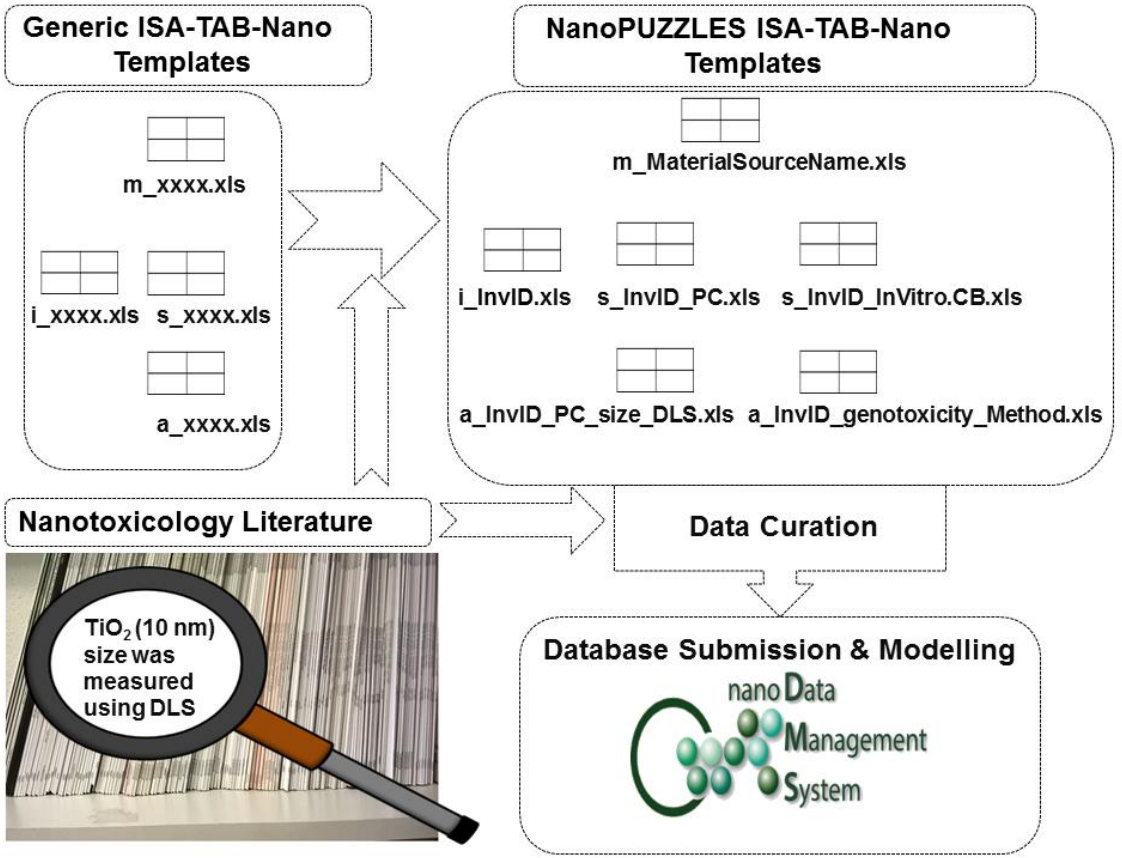
NanoPUZZLES ISA-TAB-Nano data collection: Python program[1]



1. Figure 2: Marchese Robinson, R.L. et al. *Beilstein J. Nanotechnol.*, *6*, **2015**, 1978–1999.

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

NanoPUZZLES ISA-TAB-Nano data collection: overview



- Not all assay file templates and in vivo Study file template not shown

1. Table of contents image: Marchese Robinson, R.L. et al. *Beilstein J. Nanotechnol.*, *6*, **2015**, 1978–1999.

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Challenges: stepwise sample preparation[1,2]

Study file

Protocol REF	Protocol REF	Sample Name	Factor Value [stock Sonication]	Factor Value [tested sample Sonication]
stock sample preparation	tested sample preparation	sample ID	TRUE	FALSE

OR

Protocol REF	Sample Name	Factor Value [stock Sonication]	Factor Value [tested sample Sonication]
sample preparation	sample ID	TRUE	FALSE

OR

Protocol REF	Sample Name	Factor Value [Sonication] [treatment order = 1]	Factor Value [Sonication] [treatment order = 2]
sample preparation	sample ID	TRUE	FALSE

1. Marchese Robinson, R.L. et al. *Beilstein J. Nanotechnol.*, *6*, **2015**, 1978–1999.
2. http://isatab.sourceforge.net/docs/ISA-TAB_release-candidate-1_v1.0_24nov08.pdf

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837

Challenges: temporal metadata[1]

Study file (PCCs)

Sample Name	Factor Value [medium Exposure Duration]	Unit
sample ID	5	hour

Study file (in vitro)

Sample Name	Factor Value [cells Exposure Duration]	Unit
sample ID	2	hour

- Time difference between PCCs and in vitro assay?
- Use general “Factor Value [time point]”?

1. Marchese Robinson, R.L. et al. *Beilstein J. Nanotechnol.*, *6*, **2015**, 1978–1999.

This project is funded by the European Union under the 7th Research Framework Programme [Theme: NMP-2012-SMALL-6]
Grant Agreement nr 309837