LINCS standardized small molecule information								Notes
Unique ID	LINCS Field Name	Relate to	Description	Importance (1: essential; 2: desirable / recommended; 3: optional)	Comments	Ontologies / references considered	Link to ontology / reference	Additional Notes (for development)
SM:1	SM_Name	canonical structure	The primary name for the (parent) compound (in a standardized representation) as chosen by LINCS	1	common, recognizable, name	DrugBank, PubChem, ChEMBL		
SM:2	SM_LINCS_ID	canonical structure	The global LINCS ID (parent) compound (in a standardized representaiton)	1				
SM:3	SM_AlternativeNames	canonical structure	List of synonymous compound names, drug name (if applicable), and other alternative names	1	drug names can come from common drug name available from DrugBank, TTD, NPC or other resources	DrugBank, PubChem, ChEMBL, DrugBank, TTD, NPC		
SM:4	SM_Center_Compound_ID	canonical structure	Center specific compound ID of the (parent structure) assigned by the center	1				
SM:5	SM_Center_Sample_ID	batch	Sample ID of the tested compound, referring to of the tested sample; assigned after local registry of the compound (center specific)	1	assigned by local LINCS center			
SM:6	SM_Provider	batch	Vendor or lab that supplied the compound	1	list of vendor names and other providers			
SM:7	SM_Provider_Catalog_ID	batch	ID or catalogue number assigned to the specific supplied sample by the vendor or provider	1				
SM:8	SM_Provider_Sample_ID	batch	Sample ID of the compound obtained from the provider or vendor (if available); this should include batch and / or lot number	2				
SM:9	SM_Salt	batch	Reference to counter-ions and other addends present in the compound's formulation	1	HMS LINCS has developed a three digit "salt code" used as part of the facility ID for its compounds			need a salt addend DB, which UM can provide
SM:10	SM_PubChem_CID	canonical structure	CID that corresponds to the standardized parent compound in NCBI's PubChem database; after applying the same business rules	1	This is the PubChem compound ID, and not to be confused with the substance ID.	PubChem	http://pubchem.ncbi.nlm.nih.gov/	PubChem CID of main LINCS compound representation
SM:11	SM_ChEBI_ID	canonical structure	ChEBI ID that corresponds to the standardized parent compound in NCBI's PubChem database; after applying the same business rules	2	ChEBI ID; ChEBI is curated and we may submit important structure to ChEBI	ChEBI	http://www.ebi.ac.uk/chebi/	
SM:12	SM_InChi_Parent	canonical structure	InChi representation of parent (standardized) chemical structure chemical structure generated by LINCS business rules	2	standardized small molecule representation; documented standardization protocol			
SM:13	SM_InChi_Key_Parent	canonical structure	InChi key of parent (standardized) chemical structure generated by LINCS business rules	2				derived from InChi
SM:14	SM_SMILES_Parent	canonical structure	Canonical SMILES representation of parent (standardized) chemical structure generated by LINCS business rules	1	standardized small molecule representation; documented standardization protocol			
SM:15	SM_Image_Parent	canonical structure	An image of the parent (standardized) compound produced using structure rendering software	1	standardized small molecule representation; documented standardization protocol			
SM:16	SM_InChi_Batch	batch	InChi representation of the actual sample (batch) structure	2	this includes salts and addends and no structure canonicalization; full structure			
SM:17	SM_InChi_Key_Batch	batch	InChi key of the actual sample (batch) structure	2				derived from InChi
SM:18	SM_SMILES_Batch	batch	Canonical SMILES representation of the actual sample (batch) structure	1	this includes salts and addends and no structure canonicalization; full structure			
SM:19	SM_Software		Name, Version, Link for Software used to generate InChi and SMILES	2				The software versions are available from IUPAC
SM:20	SM_Target Information_References	canonical structure	Information about the pharmacological target(s) of the compound in cells, e.g. proteins or other molecules with which the drug interacts to confer its effect. Appropriate literature references should also be provided.	2	this should include references to other databases where such information can be extracted; it appears unrealistic to maintain the available external biological activity data for LINCS compounds; if we create a database we can reference this; this will be implemented in multiple fields			it appears that this should be a separate database that links here
SM:21	SM_Molecular_Mass	batch	Molecular mass of one molecule (including addends) of the compound in Daltons (unified atomic mass units)	1	this includes salts and addends and no structure canonicalization; full structure			
SM:22	SM_Molecular_Formula	batch	String representation of the compound with addends; type and number of the different atoms in the compound; without structural details	2	this includes salts and addends and no structure canonicalization; full structure			

LINCS standardized small molecule information								Notes
Unique ID	LINCS Field Name	Relate to	Description	Importance (1: essential; 2: desirable / recommended; 3: optional)	Comments	Ontologies / references considered	Link to ontology / reference	Additional Notes (for development)
SM:23	SM_Chemical_Synthesis_Reference	batch	Reference to the synthetic procedure by which a molecule has been generated	3	suggested by MIABE; if available			
SM:24	SM_Purity	batch	Purity of the compound in percent	3	suggested by MIABE; if available			
SM:25	SM_Purity_Method	batch	Method for determining the purity of the compound	3				
SM:26	SM_Aqueous_Solubility	batch	The actual (measured) aqueous solubility of the compound in mg / mL or g / L	3	suggested by MIABE; if available			
SM:27	SM_LogP	batch	The logarithm of the actual (measured) water/octanol partition coefficient (logP) or hydrophobicity score	3	suggested by MIABE; if available			