

Why Share Biomedical Data?

- Science has always encouraged data sharing to advance.
 - It allows scientists to check and build on the work of others.
 - Saves public resources as data can be produced once and used over again by many.
- Its the right thing to do.

Why Not Share Biomedical Data?

- Some special issues exist for clinical data in regards to patient privacy issues.
- Some technological challenges exist but are not the main problem.
- “Many of the barriers to sharing and reuse are social in nature, arising from researchers’ concerns about and attitudes toward sharing their data.”

Federer *et al* (2015) <http://dx.doi.org/10.1371/journal.pone.0129506>

How to Share Data?

What is Metadata?

- **Metadata is “data that provides information about other data”.**

Example: A book entry in a library catalog is metadata about the book.

- **Multiple types of metadata**

Descriptive metadata describes individual instances of the data or the data content. Used for discovery.
(Title, abstract, author, & keywords)

What is Metadata?

- **Metadata is “data that provides information about other data”.**

Example: A book entry in a library catalog is metadata about the book.

- **Multiple types of metadata**

Descriptive metadata

Structural metadata is data about how the data is put together, the containers of data.
(File formats, database schema, versions)

What is Metadata?

- **Metadata is “data that provides information about other data”.**

Example: A book entry in a library catalog is metadata about the book.

- **Multiple types of metadata**

Descriptive metadata

Structural metadata

Administrative metadata provides information to manage a resource, such as when and how it was created, file type and other technical information, access restrictions.

What is Metadata?

- **Metadata is “data that provides information about other data”.**

Example: A book entry in a library catalog is metadata about the book.

- **Multiple types of metadata**

Descriptive metadata

Structural metadata

Administrative metadata

- **The term metadata is used differently in different communities.**

Some use it to refer to machine understandable information

Some use it only for human readable records that describe a resource

Help results

[Show quick links](#)

Filter byⁱ

◀ 1 to 25 of 109 ▶ Show 25 ⬆

UniProtKB manual ✕
(109)

Categories

[3D structure \(4\)](#)
[Cross-references \(2\)](#)
[Entry information \(5\)](#)
[Expression \(4\)](#)
[Family and domains \(9\)](#)
[Function \(16\)](#)
[Interaction \(3\)](#)
[Miscellaneous \(2\)](#)

Help Search Results

Accession

Provides a stable way of identifying UniProtKB entries

Active site

Amino acid(s) directly involved in the activity of an enzyme

Allergenic properties

Information relevant to allergenic proteins

Alternative products

Description of the different proteins generated from the same gene

Annotation score

Annotation scores provide a measure of how much annotation has been associated with a given entry or proteome.

Binary interactions

Information relevant to binary protein-protein interactions

Binding site

Binding site for any chemical group (co-enzyme, prosthetic group, etc.)

Biophysicochemical properties

Description of biophysical and physicochemical properties


```

<!-- Entry definition begins -->
- <xs:element name="entry">
  - <xs:annotation>
    <xs:documentation>Describes a UniProtKB entry.</xs:documentation>
  </xs:annotation>
- <xs:complexType>
  - <xs:sequence>
    <xs:element name="accession" type="xs:string" maxOccurs="unbounded"/>
    <xs:element name="name" type="xs:string" maxOccurs="unbounded"/>
    <xs:element name="protein" type="proteinType"/>
    <xs:element name="gene" type="geneType" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="organism" type="organismType"/>
    <xs:element name="organismHost" type="organismType" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="geneLocation" type="geneLocationType" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="reference" type="referenceType" maxOccurs="unbounded"/>
    <xs:element name="comment" type="commentType" nillable="true" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="dbReference" type="dbReferenceType" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="proteinExistence" type="proteinExistenceType"/>
    <xs:element name="keyword" type="keywordType" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="feature" type="featureType" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="evidence" type="evidenceType" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="sequence" type="sequenceType"/>
  </xs:sequence>
  - <xs:attribute name="dataset" use="required">
    - <xs:simpleType>
      - <xs:restriction base="xs:string">
        <xs:enumeration value="Swiss-Prot"/>
        <xs:enumeration value="TrEMBL"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="created" type="xs:date" use="required"/>
  <xs:attribute name="modified" type="xs:date" use="required"/>
  <xs:attribute name="version" type="xs:int" use="required"/>
</xs:complexType>
</xs:element>
<!-- Entry definition ends -->

```


core:Annotation (rdf:type owl:Class)

rdfs:comment	Description of a resource on a specific topic. xsd:string
--------------	---

core:Attribution (rdf:type owl:Class)

rdfs:comment	Entity used to attach evidence or provenance to a rdf statement via reification. @en
--------------	--

core:Beta_Strand_Annotation (rdf:type owl:Class)

rdfs:subClassOf	core:Secondary_Structure_Annotation
rdfs:comment	Beta strand regions within the experimentally determined protein structure xsd:string
rdfs:label	Strand xsd:string
rdfs:seeAlso	http://www.uniprot.org/manual/strand

core:Binding_Site_Annotation (rdf:type owl:Class)

rdfs:subClassOf	core:Site_Annotation
rdfs:comment	Binding site for any chemical group (co-enzyme, prosthetic group, etc.). xsd:string
rdfs:label	Binding Site xsd:string
rdfs:seeAlso	http://www.uniprot.org/manual/binding

core:Biophysicochemical_Annotation (rdf:type owl:Class)

rdfs:subClassOf	core:Annotation
rdfs:comment	Biophysical and physicochemical data such as pH dependence, temperature dependence, kinetic parameters, redox potentials, and maximal absorption. xsd:string
rdfs:label	Biophysicochemical Property xsd:string
rdfs:seeAlso	core:biophysicochemical_properties

[NCBI](#) > [GEO](#) > [Accession Display](#) 

Scope: Format: Amount: GEO accession:

Series GSE54293

Query DataSets for GSE54293

Status	Public on Jan 23, 2014
Title	Akt inhibitor MK2206 prevents influenza A(H1N1)pdm09 virus infection in vitro
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	The influenza A(H1N1)pdm09 virus caused a global flu pandemic in 2009 and contributes to seasonal epidemics. Different treatment and prevention options for influenza have been developed and applied with limited success. Here we report that an Akt inhibitor MK2206 possesses potent antiviral activity against influenza A(H1N1)pdm09 virus in vitro. We showed that MK2206 blocks the entry of different A(H1N1)pdm09 strains into cells. Moreover, MK2206 prevented A(H1N1)pdm09-mediated activation of cellular signaling pathways and the development of cellular immune responses. Importantly, A(H1N1)pdm09 virus was unable to develop resistance to MK2206. Thus, MK2206 is a potent anti-influenza A(H1N1)pdm09 agent.

Overall design	Total RNA obtained from NCI-H1666 cells, which are non-small cell lung cancer cell line. NCI-H1666 cells were non- or MK2206-treated (10 μ M) and mock- or virus-infected (A/Helsinki/p14/2009) at moi of 3.
----------------	--

Contributor(s) Denisova OV, Virtanen S, VonSchantz C, Bychkov D, Desloovere J, Soderholm S, Theisen LL, Tynell J, Ikonen N, Vashchinkina E, Nyman TA, Matikainen S, Kallioniemi O, Julkunen I, Muller CP, Saelens X, Verkhusha VV, Kainov DE

Citation(s) Denisova OV, Söderholm S, Virtanen S, Von Schantz C et al. Akt inhibitor MK2206 prevents influenza pH1N1 virus infection in vitro. *Antimicrob Agents Chemother* 2014 Jul;58(7):3689-96. PMID: [24752266](#)

Submission date	Jan 22, 2014
Last update date	Aug 30, 2016
Contact name	Denis Kainov
E-mail	denis.kainov@helsinki.fi
Organization name	Helsinki University
Department	Institute for Molecular M

```
<Series iid="GSE54293">
  <Status database="GEO">
    <Submission-Date>2014-01-22</Submission-Date>
    <Release-Date>2014-01-23</Release-Date>
    <Last-Update-Date>2016-08-30</Last-Update-Date>
  </Status>
  <Title>Akt inhibitor MK2206 prevents influenza A(H1N1)pdm09 virus infection in vitro</Title>
  <Accession database="GEO">GSE54293</Accession>
  <Pubmed-ID>24752266</Pubmed-ID>
  <Summary>
```

The influenza A(H1N1)pdm09 virus caused a global flu pandemic in 2009 and contributes to seasonal epidemics. Different treatment and prevention options for influenza have been developed and applied with limited success. Here we report that an Akt inhibitor MK2206 possesses potent antiviral activity against influenza A(H1N1)pdm09 virus in vitro. We showed that MK2206 blocks the entry of different A(H1N1)pdm09 strains into cells. Moreover, MK2206 prevented A(H1N1)pdm09-mediated activation of cellular signaling pathways and the development of cellular immune responses. Importantly, A(H1N1)pdm09 virus was unable to develop resistance to MK2206. Thus, MK2206 is a potent anti-influenza A(H1N1)pdm09 agent.

<Overall-Design>
Total RNA obtained from NCI-H1666 cells, which are non-small cell lung cancer cell line. NCI-H1666 cells were non- or MK2206-treated (10 μ M) and mock- or virus-infected (A/Helsinki/p14/2009) at moi of 3.

```
</Overall-Design>
<Type>Expression profiling by array</Type>
<Contributor-Ref ref="contrib3" position="1" />
<Contributor-Ref ref="contrib4" position="2" />
<Contributor-Ref ref="contrib5" position="3" />
```


Biological Data Standards

- **Why do we need Standards?**

- Without standards it becomes impossible to reuse or combine similar data types. Data is not interoperable.
- There are two types of interoperability

- **Syntactic interoperability**

means two or more systems are capable of communicating and exchanging data. Specific data formats and protocols are required.

Biological Data Standards

- **Why do we need Standards?**

- Without standards it becomes impossible to reuse or combine similar data types. Data is not interoperable.

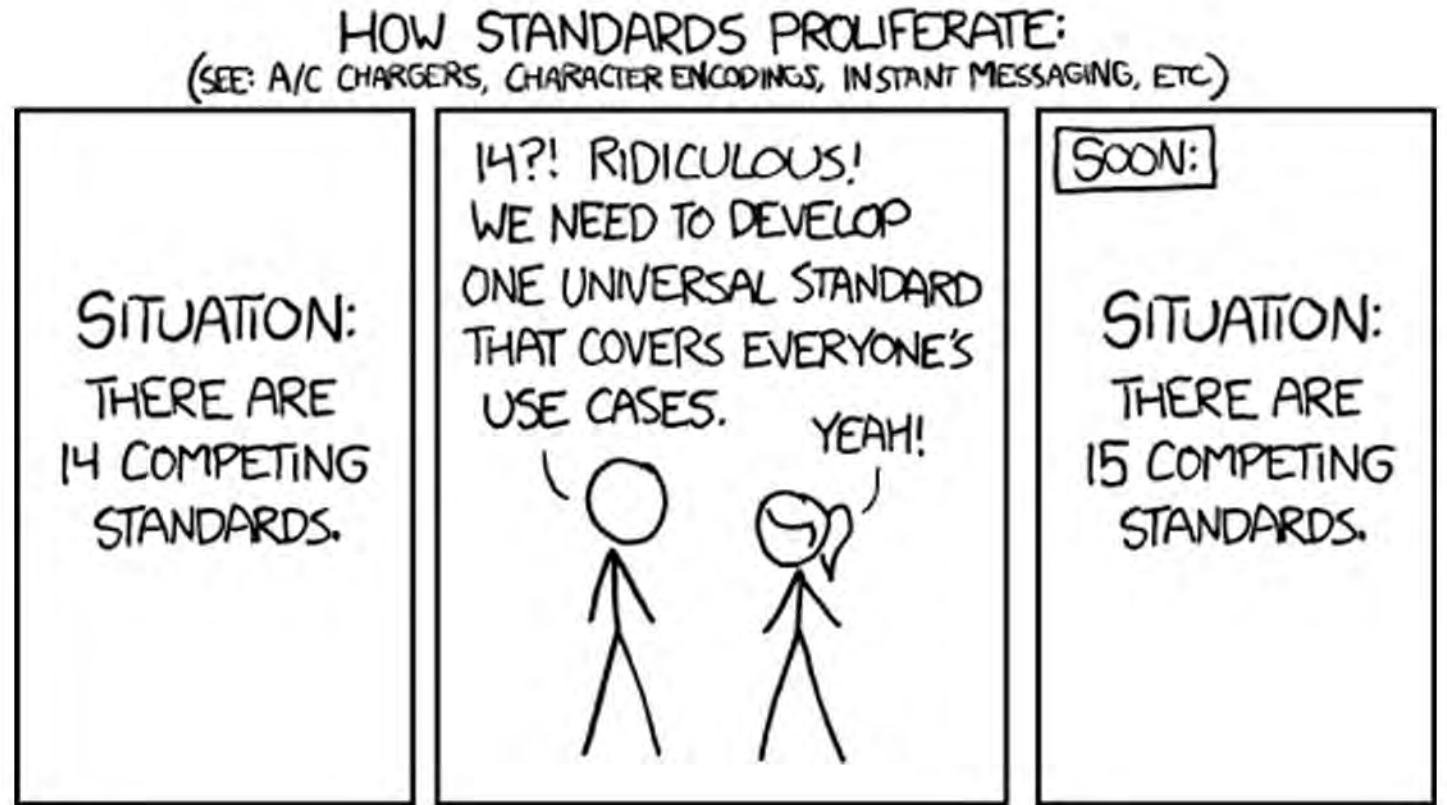
- There are two types of interoperability

- **Syntactic interoperability**

- **Semantic interoperability**

is the ability to automatically interpret the information exchanged meaningfully and accurately to produce a useful exchange for the end users of both systems. Specific data models including vocabularies and/or ontologies are required to make sure both systems are communicating.

What Standards to Use?



<http://imgs.xkcd.com/comics/standards.png>

Minimum Information Standards

- The minimum information standard is a set of guidelines for reporting data derived by relevant methods in biosciences.
 - What is good about it?
As a minimal standard it tends to be **simpler to meet and create** than a complete standard.
 - What is not good?
It is a **minimal** standard and as such not as rich as it could be and may not meet all usecases.

Look for Standards at Biosharing.org

The screenshot shows the biosharing.org website. At the top left is the logo with the text "biosharing.org" and "standards, databases, policies" below it, and the "elixir" logo with "UNITED KINGDOM" below it. To the right of the logo is a navigation bar with buttons for "Standards", "Databases", "Policies", "Collections", "Add/Claim Content", "Stats", and "Log in or Register". Below the navigation bar is a main heading: "A curated, informative and educational resource on inter-related data standards, databases, and policies in the life, environmental and biomedical sciences". Under this heading are three white boxes. The first box is titled "Find" and contains a thumbs-up icon and the text "Recommendations", with a description: "Standards and/or databases recommended by journal or funder data policies." The second box is titled "Discover" and contains a network icon and the text "Collections", with a description: "Standards and/or databases grouped by domain, species or organization." The third box is titled "Learn" and contains a graduation cap icon and the text "Educational", with a description: "About standards, their use in databases and policies, and how we can help you." At the bottom of the page is a dark blue footer. On the left is a search bar with the placeholder "Search BioSharing" and a "Search" button. Below the search bar are four checkboxes, all of which are checked: "Standards", "Databases", "Policies", and "Collections/Recommendations". On the right side of the footer are two sections: "Advanced Search" with a magnifying glass icon and the text "Fine grained control over your search.", and "Search Wizard" with a speech bubble icon and the text "Let us guide you to your results."

biosharing.org standards, databases, policies **elixir** UNITED KINGDOM

Standards Databases Policies Collections Add/Claim Content Stats Log in or Register

A curated, informative and educational resource on inter-related data standards, databases, and policies in the life, environmental and biomedical sciences

Find

Recommendations

Standards and/or databases recommended by journal or funder data policies.

Discover

Collections

Standards and/or databases grouped by domain, species or organization.

Learn

Educational

About standards, their use in databases and policies, and how we can help you.

Search BioSharing Search

☒ Standards ☒ Databases ☒ Policies ☒ Collections/Recommendations

Advanced Search

Fine grained control over your search.

Search Wizard

ask biosharing

Let us guide you to your results.

Standards at Biosharing.org

Registry	Name	Abbreviation	Type	Domain	Taxonomy	Related Database	Related Standard	Related Policy	In Collection/Recommendation
	CDISC Analysis Data Model	CDISC ADaM	Standard	<div>Analysis</div> <div>Biomedical Science</div> <div>Clinical Trial</div> <div>Data Model</div> <div>Data Transformation</div>	<div>Homo Sapiens</div>	None	<div>CDISC SEND</div> <div>CDISC SDTM</div> <div>CDISC CDASH</div> <div>CDISC Define.xml</div> <div>CDISC ODM</div> <div>Plus 6 more...</div>	None	<div>eTRIKS Standards Starter Pack</div> <div>Clinical Research (CDISC)</div>
	mz peptide and protein Identification Markup Language	mzIdentML	Standard	<div>Centrally Registered Identifier</div> <div>Identification</div> <div>Life Science</div> <div>Protein</div> <div>Standard</div>	<div>All</div>	None	<div>mzQuantML</div> <div>PSI-MS CV</div> <div>PSI SpML</div> <div>QCML</div>	None	None
	mz Quantitative Markup Language	mzQuantML	Standard	<div>Life Science</div> <div>Protein</div> <div>Quantification</div> <div>Quantity</div>	<div>All</div>	PRIDE	<div>PSI-MS CV</div> <div>mzIdentML</div> <div>mzML</div> <div>PSI SpML</div> <div>HUPO-PSI TraML</div> <div>Plus 2 more...</div>	None	None
	Anatomical Entity Ontology	AEO	Standard	<div>Anatomy</div> <div>Life Science</div>	<div>Invertebrata</div> <div>Vertebrata</div>	None	<div>CARO</div> <div>EHDA2</div>	None	<div>OBO Foundry</div>
	Biological SPatial Ontology	BSPO	Standard	<div>Anatomy</div> <div>Life Science</div>	<div>Animalia</div> <div>Fungi</div> <div>Viridiplantae</div>	ZFIN	<div>CARO</div>	None	<div>OBO Foundry</div>
	Common Anatomy Reference Ontology	CARO	Standard	<div>Anatomy</div> <div>Life Science</div> <div>Ontology Standardization</div>	<div>Vertebrata</div>	None	<div>AEO</div> <div>BSPO</div> <div>VSAO</div> <div>FMA</div>	None	<div>OBO Foundry</div>