

# ISO 20691:2022-11 (E)

## Biotechnology - Requirements for data formatting and description in the life sciences

<b>Contents</b>		<b>Page</b>
Foreword .....		v
Introduction .....		vi
1	Scope .....	1
2	Normative references .....	1
3	Terms and definitions .....	1
4	Recommendations and requirements for the description of entities and concepts in life science data .....	8
4.1	General .....	8
4.2	Recommended ubiquitous identifier scheme for biological and conceptual entities .....	8
4.2.1	URI provisions .....	8
4.2.2	IRI provisions .....	9
4.2.3	Relationship between URI and IRI .....	10
4.3	Formatting data and contextual descriptive data (metadata) for biological entities and concepts .....	10
4.3.1	General .....	10
4.3.2	Version control .....	10
4.3.3	Arbitrary Limits .....	10
4.3.4	Character sets .....	10
4.3.5	Machine readability .....	10
4.3.6	Knowledge representation .....	11
5	Technical and organizational recommendations and requirements for data formats .....	11
5.1	General .....	11
5.2	Organizational responsibilities .....	11
5.3	Documentation .....	12
5.4	Versioning and change log .....	12
5.5	Compatibility .....	12
5.6	Extensibility .....	12
5.7	Compression .....	12
5.8	Structural and control elements .....	12
5.9	Requirements for data types within formats .....	13
5.9.1	General .....	13
5.9.2	Encoding of numerical quantity values .....	13
5.9.3	Encoding of character strings .....	13
5.9.4	Encoding of sequence data .....	13
5.9.5	Time data .....	13
5.9.6	Boolean data .....	13
5.9.7	Biological Imaging data .....	14
5.10	Consistency and compatibility .....	14
5.11	Data integrity .....	14
5.12	Format validation .....	14
5.13	Data provenance .....	14
6	Semantic recommendations and requirements for data formats .....	15
6.1	General .....	15
6.2	Minimum consensus information for annotation of biological data .....	15
6.2.1	General .....	15

6.2.2	Species .....	16
6.2.3	Sex .....	16
6.2.4	Age .....	16
6.2.5	Organ .....	16
6.2.6	Tissue .....	16
6.2.7	Cell type .....	16
6.2.8	Identifiable objects .....	16
ISO 20691:2022(E) 6.2.9	Identifiable processes .....	17
6.2.10	Manipulated entities .....	17
6.2.11	Analytical, experimental and computational technology .....	17
6.2.12	Biological or analytical question .....	17
6.2.13	Technology-specific data .....	17
6.3	Syntax and reification .....	19
7	<b>Requirements for terminologies and ontologies suitable for annotation of biological data</b> .....	19
7.1	General .....	19
7.2	<b>Requirements for biological ontologies</b> .....	19
7.2.1	Maintainer .....	19
7.2.2	Maintenance of the ontology .....	19
7.2.3	Ontology syntax .....	20
7.2.4	Linking to other ontologies and term reuse .....	20
7.2.5	Licensing and attribution .....	20
7.2.6	Stable URIs and versioning information .....	20
7.2.7	Community involvement .....	20
7.2.8	Language .....	20
8	<b>Requirements for domain specific data standards</b> .....	20
8.1	General .....	20
8.2	<b>Specific requirements for domain specific data standards</b> .....	20
8.2.1	Maintainer .....	20
8.2.2	Maintenance of the data standard .....	21
8.2.3	Data standard syntax .....	21
8.2.4	Linking to other data standards .....	21
8.2.5	Licensing and attribution .....	21
8.2.6	Stable URIs and versioning information .....	21
8.2.7	Community involvement .....	21
8.2.8	Language .....	21
9	<b>Requirements for data repositories for biological data</b> .....	22
9.1	General .....	22
9.2	<b>Requirements for data repositories of biological data</b> .....	22
9.2.1	Maintainer .....	22
9.2.2	Maintenance of the repository .....	22
9.2.3	Repository structure .....	22
9.2.4	Linking to other repositories .....	22
9.2.5	Licensing and attribution .....	22
9.2.6	Stable URIs and versioning information .....	22
9.2.7	Data visibility .....	23
9.2.8	Community involvement .....	23
9.2.9	Language .....	23
Annex A (informative)	Examples of common formats for life science data .....	24
Annex B (informative)	Minimum reporting standards for data, models and metadata .....	37
Bibliography	.....	47