**Minimum Information about any (x) Sequence (MIxS) packages (version 5.0)**

MIxS environmental package provides the list of descriptors / parameters to describe a biosample and other related information. The extensive range of descriptors are contextual data for accurate and consistent description of the sequenced sample, its environment and sequencing experiments performed with the sample. There are 17 different MIxS packages (version 5.0) currently available.

MIxS package descriptors have been provided in five different sections namely, (i) environment, (ii) investigation, (iii) mixs extension (one of the 17 MIxS biosample type: ex: soil, water, sediment, etc), (iv) nucleic acid sequence source and (v) sequencing.

Though each of the 17 MIxS packages has a long list of descriptors, only some of the descriptors are indicated as **mandatory** to be provided for the biosample. Among the mandatory descriptors, ten are shared across all 17 packages.

**Shared mandatory descriptors in all of the 17 MIxS packages are:**

1. submitted\_to\_insdc (Depending on the study (large-scale e.g. done with next generation sequencing technology, or small-scale) sequences have to be submitted to SRA (Sequence Read Archive), DRA (DDBJ Read Archive) or via the classical Webin/Sequin systems to Genbank, ENA and DDBJ)
2. investigation\_type (Nucleic Acid Sequence Report is the root element of all MIGS/MIMS compliant reports as standardized by Genomic Standards Consortium. This field is either eukaryote, bacteria, virus, plasmid, organelle, metagenome, mimarks-survey, mimarks-specimen, metatranscriptome, single amplified genome, metagenome-assembled genome, or uncultivated viral genome)
3. project\_name (Name of the project within which the sequencing was organized)
4. lat\_lon (The geographical origin of the sample as defined by latitude and longitude)
5. geo\_loc\_name (The geographical origin of the sample as defined by the country or sea name followed by specific region name)
6. collection\_date (The time of sampling, either as an instance (single point in time) or interval. In case no exact time is available, the date/time can be right truncated i.e. all of these are valid times: 2008-01-23T19:23:10+00:00; 2008-01-23T19:23:10; 2008-01-23; 2008-01; 2008; Except: 2008-01; 2008 all are ISO8601 compliant)
7. env\_broad\_scale (broad-scale environmental context).
8. env\_local\_scale (local environmental context)
9. env\_medium (environmental medium)
10. seq\_meth (Sequencing method used; e.g. Sanger, pyrosequencing, ABI-solid)

**Unique descriptors in the individual MIxS packages:**

There are some descriptors present in only one of the given packages (unique) when compared with other 16 packages. However, unique descriptors in a specific MIxS package are not necessarily mandatory descriptors. We have provided unique descriptors along with their definitions of each MIxS package in this document for a quick reference. *Further, in the respective FICUS (Facilities Integrating Collaborations for User Science) study / proposal metadata spreadsheet, each MIxS term is provided (row number 3) along with a short description in row # 4 and the type of data required in row # 5.*

It should also be noted that in two MIxS packages, i) Microbial mat/biofilm, ii) Miscellaneous natural or artificial environment, there are no unique descriptors present. The differences in the descriptors among various MIxS packages would help us while assigning the biosample to the specific package.

We have mapped the MIxS package specific descriptors to GOLD (Genomes Online Database) fields. The metadata file specific to the ‘FICUS Proposal / Study’ contains GOLD fields (MIxS descriptors) for which data are required from the PI and the fields that are already available in the GOLD database that PI may review / update.

**1. MIxS Air Package:**

The MIxS air package provides a list of descriptors for biosample that are associated with atmosphere / air including indoor air or air filter (air metagenome), ground-level atmosphere (dust metagenome), laminar flow hood or man-made tunnel (aerosol metagenome).

In addition to the ten shared mandatory descriptors given in **page 1**, “**altitude**” is also a **mandatory** field in the air package.

*(****Altitude:*** *Altitude is a term used to identify heights of objects such as airplanes, space shuttles, rockets, atmospheric balloons and heights of places such as atmospheric layers and clouds. It is used to measure the height of an object which is above the earth’s surface. In this context, the altitude measurement is the vertical distance between the earth's surface above sea level and the sampled position in the air).*

The following 11 descriptors are **unique** in the MIxS air package (that are not present in other 16 MIxS packages):

1. barometric\_press (Force per unit area exerted against a surface by the weight of air above that surface)
2. carb\_monoxide (Carbon monoxide (gas) amount or concentration at the time of sampling)
3. humidity (Amount of water vapour in the air, at the time of sampling)
4. oxygen (Oxygen (gas) amount or concentration at the time of sampling)
5. pollutants (Pollutant types and, amount or concentrations measured at the time of sampling; can report multiple pollutants by entering numeric values preceded by name of pollutant)
6. resp\_part\_matter (Concentration of substances that remain suspended in the air, and comprise mixtures of organic and inorganic substances (PM10 and PM2.5); can report multiple PM's by entering numeric values preceded by name of PM)
7. solar\_irradiance (The amount of solar energy that arrives at a specific area of a surface during a specific time interval)
8. ventilation\_rate (Ventilation rate of the system in the sampled premises)
9. volatile\_org\_comp (Concentration of carbon-based chemicals that easily evaporate at room temperature; can report multiple volatile organic compounds by entering numeric values preceded by name of compound)
10. wind\_direction (Wind direction is the direction from which a wind originates)
11. wind\_speed (Speed of wind measured at the time of sampling)

An example biosample that is using MIxS Air package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/SAMN02904796>

**2. MIxS Built environment package**

The MIxS-built environment package provides the list of descriptors for biosample that are associated with various built environments including hospitals, offices, factories, bathroom, buildings, factory indoor.

In addition to the ten shared mandatory descriptors given in **page 1**, another 16 descriptors are provided as **mandatory** for this package:

1. rel\_air\_humidity (Partial vapor and air pressure, density of the vapor and air, or by the actual mass of the vapor and air)
2. abs\_air\_humidity (Actual mass of water vapor - mh20 - present in the air water vapor mixture)
3. air\_temp (Temperature of the air at the time of sampling)
4. build\_occup\_type (The primary function for which a building or discrete part of a building is intended to be used)
5. carb\_dioxide (Carbon dioxide (gas) amount or concentration at the time of sampling)
6. ventilation\_type (Ventilation system used in the sampled premises)
7. organism\_count (Total cell count of any organism (or group of organisms) per gram, volume or area of sample, should include name of organism followed by count. The method that was used for the enumeration (e.g. qPCR, atp, mpn, etc.) Should also be provided. (example: total prokaryotes; 3.5e7 cells per ml; qpcr))
8. indoor\_space (A distinguishable space within a structure, the purpose for which discrete areas of a building is used)
9. filter\_type (A device which removes solid particulates or airborne molecular contaminants)
10. heat\_cool\_type (Methods of conditioning or heating a room or building)
11. building\_setting (A location (geography) where a building is set)
12. light\_type (Application of light to achieve some practical or aesthetic effect. Lighting includes the use of both artificial light sources such as lamps and light fixtures, as well as natural illumination by capturing daylight. Can also include absence of light)
13. space\_typ\_state (Customary or normal state of the space)
14. typ\_occup\_density (Customary or normal density of occupants)
15. occup\_samp (Number of occupants present at time of sample within the given space)
16. occup\_density\_samp (Average number of occupants at time of sampling per square footage)

Among the 253 descriptors given in the MIxS-built environment package, the following 157 are **unique** to it:

1. abs\_air\_humidity (Actual mass of water vapor - mh20 - present in the air water vapor mixture)
2. address (The street name and building number where the sampling occurred.)
3. adj\_room (List of rooms (room number, room name) immediately adjacent to the sampling room)
4. aero\_struc (Aerospace structures typically consist of thin plates with stiffeners for the external surfaces, bulkheads and frames to support the shape and fasteners such as welds, rivets, screws and bolts to hold the components together)
5. air\_temp (Temperature of the air at the time of sampling)
6. amount\_light (The unit of illuminance and luminous emittance, measuring luminous flux per unit area)
7. arch\_struc (An architectural structure is a human-made, free-standing, immobile outdoor construction )
8. avg\_dew\_point (The average of dew point measures taken at the beginning of every hour over a 24 hour period on the sampling day)
9. avg\_occup (Daily average occupancy of room)
10. avg\_temp (The average of temperatures taken at the beginning of every hour over a 24 hour period on the sampling day)
11. bathroom\_count (The number of bathrooms in the building)
12. bedroom\_count (The number of bedrooms in the building)
13. build\_docs (The building design, construction and operation documents)
14. build\_occup\_type (The primary function for which a building or discrete part of a building is intended to be used)
15. building\_setting (A location (geography) where a building is set)
16. built\_struc\_age (The age of the built structure since construction)
17. built\_struc\_set (The characterization of the location of the built structure as high or low human density)
18. built\_struc\_type (A physical structure that is a body or assemblage of bodies in space to form a system capable of supporting loads)
19. ceil\_area (The area of the ceiling space within the room)
20. ceil\_cond (The physical condition of the ceiling at the time of sampling; photos or video preferred; use drawings to indicate location of damaged areas)
21. ceil\_finish\_mat (The type of material used to finish a ceiling)
22. ceil\_struc (The construction format of the ceiling)
23. ceil\_texture (The feel, appearance, or consistency of a ceiling surface)
24. ceil\_thermal\_mass (The ability of the ceiling to provide inertia against temperature fluctuations. Generally this means concrete that is exposed. A metal deck that supports a concrete slab will act thermally as long as it is exposed to room air flow)
25. ceil\_type (The type of ceiling according to the ceiling's appearance or construction)
26. ceil\_water\_mold (Signs of the presence of mold or mildew on the ceiling)
27. cool\_syst\_id (The cooling system identifier)
28. date\_last\_rain (The date of the last time it rained)
29. dew\_point (The temperature to which a given parcel of humid air must be cooled, at constant barometric pressure, for water vapor to condense into water. )
30. door\_comp\_type (The composite type of the door)
31. door\_cond (The phsical condition of the door)
32. door\_direct (The direction the door opens)
33. door\_loc (The relative location of the door in the room)
34. door\_mat (The material the door is composed of)
35. door\_move (The type of movement of the door)
36. door\_size (The size of the door)
37. door\_type (The type of door material)
38. door\_type\_metal (The type of metal door)
39. door\_type\_wood (The type of wood door)
40. door\_water\_mold (Signs of the presence of mold or mildew on a door)
41. drawings (The buildings architectural drawings; if design is chosen, indicate phase-conceptual, schematic, design development, and construction documents)
42. elevator (The number of elevators within the built structure)
43. escalator (The number of escalators within the built structure)
44. exp\_duct (The amount of exposed ductwork in the room)
45. exp\_pipe (The number of exposed pipes in the room)
46. ext\_door (The number of exterior doors in the built structure)
47. ext\_wall\_orient (The orientation of the exterior wall)
48. ext\_window\_orient (The compass direction the exterior window of the room is facing)
49. filter\_type (A device which removes solid particulates or airborne molecular contaminants)
50. fireplace\_type (A firebox with chimney)
51. floor\_age (The time period since installment of the carpet or flooring)
52. floor\_area (The area of the floor space within the room)
53. floor\_cond (The physical condition of the floor at the time of sampling; photos or video preferred; use drawings to indicate location of damaged areas)
54. floor\_count (The number of floors in the building, including basements and mechanical penthouse)
55. floor\_finish\_mat (The floor covering type; the finished surface that is walked on)
56. floor\_struc (Refers to the structural elements and subfloor upon which the finish flooring is installed)
57. floor\_thermal\_mass (The ability of the floor to provide inertia against temperature fluctuations)
58. floor\_water\_mold (Signs of the presence of mold or mildew in a room)
59. freq\_clean (The number of times the building is cleaned per week)
60. freq\_cook (The number of times a meal is cooked per week)
61. furniture (The types of furniture present in the sampled room)
62. gender\_restroom (The gender type of the restroom)
63. hall\_count (The total count of hallways and cooridors in the built structure)
64. handidness (The handidness of the individual sampled)
65. heat\_cool\_type (Methods of conditioning or heating a room or building)
66. heat\_deliv\_loc (The location of heat delivery within the room)
67. heat\_system\_deliv\_meth (The method by which the heat is delivered through the system)
68. heat\_system\_id (The heating system identifier)
69. height\_carper\_fiber (The average carpet fiber height in the indoor environment)
70. indoor\_space (A distinguishable space within a structure, the purpose for which discrete areas of a building is used)
71. indoor\_surf (Type of indoor surface)
72. inside\_lux (The recorded value at sampling time (power density))
73. int\_wall\_cond (The physical condition of the wall at the time of sampling; photos or video preferred; use drawings to indicate location of damaged areas)
74. last\_clean (The last time the floor was cleaned (swept, mopped, vacuumed))
75. light\_type (Application of light to achieve some practical or aesthetic effect. Lighting includes the use of both artificial light sources such as lamps and light fixtures, as well as natural illumination by capturing daylight. Can also include absence of light)
76. max\_occup (The maximum amount of people allowed in the indoor environment)
77. mech\_struc (mechanical structure: a moving structure)
78. number\_pets (The number of pets residing in the sampled space)
79. number\_plants (The number of plant(s) in the sampling space)
80. number\_resident (The number of individuals currently occupying in the sampling location)
81. occup\_density\_samp (Average number of occupants at time of sampling per square footage)
82. occup\_document (The type of documentation of occupancy)
83. occup\_samp (Number of occupants present at time of sample within the given space)
84. pres\_animal (The number and type of animals present in the sampling space)
85. quad\_pos (The quadrant position of the sampling room within the building)
86. rel\_air\_humidity (Partial vapor and air pressure, density of the vapor and air, or by the actual mass of the vapor and air)
87. rel\_humidity\_out (The recorded outside relative humidity value at the time of sampling)
88. rel\_samp\_loc (The sampling location within the train car)
89. room\_air\_exch\_rate (The rate at which outside air replaces indoor air in a given space)
90. room\_architec\_element (The unique details and component parts that, together, form the architecture of a distinguisahable space within a built structure)
91. room\_condt (The condition of the room at the time of sampling)
92. room\_connected (List of rooms connected to the sampling room by a doorway)
93. room\_count (The total count of rooms in the built structure including all room types)
94. room\_dim (The length, width and height of sampling room)
95. room\_door\_dist (Distance between doors (meters) in the hallway between the sampling room and adjacent rooms)
96. room\_door\_share (List of room(s) (room number, room name) sharing a door with the sampling room)
97. room\_hallway (List of room(s) (room number, room name) located in the same hallway as sampling room)
98. room\_loc (The position of the room within the building)
99. room\_moist\_damage\_hist (The history of moisture damage or mold in the past 12 months. Number of events of moisture damage or mold observed)
100. room\_net\_area (The net floor area of sampling room. Net area excludes wall thicknesses)
101. room\_occup (Count of room occupancy at time of sampling)
102. room\_samp\_pos (The horizontal sampling position in the room relative to architectural elements)
103. room\_type (The main purpose or activity of the sampling room. A room is any distinguishable space within a structure)
104. room\_vol (Volume of sampling room)
105. room\_wall\_share (List of room(s) (room number, room name) sharing a wall with the sampling room)
106. room\_window\_count (Number of windows in the room)
107. samp\_floor (The floor of the building, where the sampling room is located)
108. samp\_room\_id (Sampling room number. This ID should be consistent with the designations on the building floor plans)
109. samp\_sort\_meth (Method by which samples are sorted; open face filter collecting total suspended particles, prefilter to remove particles larger than X micrometers in diameter, where common values of X would be 10 and 2.5 full size sorting in a cascade impactor.)
110. samp\_time\_out (The recent and long term history of outside sampling)
111. samp\_weather (The weather on the sampling day)
112. season (The season when sampling occurred)
113. season\_use (The seasons the space is occupied)
114. shading\_device\_cond (The physical condition of the shading device at the time of sampling)
115. shading\_device\_loc (The location of the shading device in relation to the built structure)
116. shading\_device\_mat (The material the shading device is composed of)
117. shading\_device\_type (The type of shading device)
118. shading\_device\_water\_mold (Signs of the presence of mold or mildew on the shading device)
119. space\_typ\_state (Customary or normal state of the space)
120. specific (The building specifications. If design is chosen, indicate phase: conceptual, schematic, design development, construction documents)
121. specific\_humidity (The mass of water vapour in a unit mass of moist air, usually expressed as grams of vapour per kilogram of air, or, in air conditioning, as grains per pound.)
122. substructure\_type (The substructure or under building is that largely hidden section of the building which is built off the foundations to the ground floor level)
123. surf\_air\_cont (Contaminant identified on surface)
124. surf\_humidity (Surfaces: water activity as a function of air and material moisture)
125. surf\_material (Surface materials at the point of sampling)
126. surf\_moisture (Water held on a surface)
127. surf\_moisture\_ph (ph measurement of surface)
128. surf\_temp (Temperature of the surface at the time of sampling)
129. temp\_out (The recorded temperature value at sampling time outside)
130. train\_line (The subway line name)
131. train\_stat\_loc (The train station collection location)
132. train\_stop\_loc (The train stop collection location)
133. typ\_occup\_density (Customary or normal density of occupants)
134. vis\_media (The building visual media)
135. wall\_area (The total area of the sampled room's walls)
136. wall\_const\_type (The building class of the wall defined by the composition of the building elements and fire-resistance rating.)
137. wall\_finish\_mat (The material utilized to finish the outer most layer of the wall )
138. wall\_height (The average height of the walls in the sampled room)
139. wall\_loc (The relative location of the wall within the room)
140. wall\_surf\_treatment (The surface treatment of interior wall)
141. wall\_texture (The feel, appearance, or consistency of a wall surface)
142. wall\_thermal\_mass (The ability of the wall to provide inertia against temperature fluctuations. Generally this means concrete or concrete block that is either exposed or covered only with paint)
143. wall\_water\_mold (Signs of the presence of mold or mildew on a wall)
144. water\_feat\_size (The size of the water feature)
145. water\_feat\_type (The type of water feature present within the building being sampled)
146. weekday (The day of the week when sampling occurred)
147. window\_cond (The physical condition of the window at the time of sampling)
148. window\_cover (The type of window covering)
149. window\_horiz\_pos (The horizontal position of the window on the wall)
150. window\_loc (The relative location of the window within the room)
151. window\_mat (The type of material used to finish a window)
152. window\_open\_freq (The number of times windows are opened per week)
153. window\_size (The window's length and width)
154. window\_status (Defines whether the windows were open or closed during environmental testing)
155. window\_type (The type of windows)
156. window\_vert\_pos (The vertical position of the window on the wall)
157. window\_water\_mold (Signs of the presence of mold or mildew on the window.)

Example biosample that is using the MIxS-Built environment package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/13911795>

<https://www.ncbi.nlm.nih.gov/biosample/10869573>

**3. MIxS Host-associated package**

MIxS Host-associated package provides the list of descriptors that are associated with the host (in general); please note that there are other *specific* host-associated MIxS packages available and refer elsewhere in this document whether they are applicable to the biosample studied before assigning the biosample to the MIxS host-associated package.

The **mandatory** fields for the host-associated package are the same ten given in shared mandatory descriptors provided in **page 1**.

The host-associated package contains a total of 141 descriptors of which, the following eight are **unique** to it:

1. gravidity (Whether or not subject is gravid, and if yes date due or date post-conception, specifying which is used)
2. host\_blood\_press\_diast (Resting diastolic blood pressure, measured as mm mercury)
3. host\_blood\_press\_syst (Resting systolic blood pressure, measured as mm mercury)
4. host\_body\_habitat (Original body habitat where the sample was obtained from)
5. host\_color (The color of host)
6. host\_growth\_cond (Literature reference giving growth conditions of the host)
7. host\_shape (Morphological shape of host )
8. host\_substrate (The growth substrate of the host )

Example biosample that is using the MIxS Host-associated package can be accessed from here: <https://www.ncbi.nlm.nih.gov/biosample/SAMN02232114>

**4. MIxS Human-associated package**

The MIxS human-associated package provide a list of 143 descriptors to describe the human-associated sample used for sequencing; this package can be selected when the sample belongs to human but not from the human organ systems - gut / oral / skin / vaginal, since they have separate MIxS packages.

The **mandatory** descriptors of the human-associated package are the same ten as given in the shared mandatory descriptors in **page 1** of this document.

Further, the following 19 descriptors are **unique** to the human-associated package:

1. amniotic\_fluid\_color (Specification of the color of the amniotic fluid sample)
2. blood\_blood\_disord (History of blood disorders; can include multiple disorders)
3. diet\_last\_six\_month (Specification of major diet changes in the last six months, if yes the change should be specified)
4. drug\_usage (Any drug used by subject and the frequency of usage; can include multiple drugs used)
5. foetal\_health\_stat (Specification of foetal health status, should also include abortion)
6. gestation\_state (Specification of the gestation state)
7. host\_hiv\_stat (HIV status of subject, if yes HAART initiation status should also be indicated as [YES or NO])
8. kidney\_disord (History of kidney disorders; can include multiple disorders)
9. maternal\_health\_stat (Specification of the maternal health status)
10. nose\_throat\_disord (History of nose-throat disorders; can include multiple disorders)
11. pet\_farm\_animal (Specification of presence of pets or farm animals in the environment of subject, if yes the animals should be specified; can include multiple animals present)
12. pulmonary\_disord (History of pulmonary disorders; can include multiple disorders)
13. smoker (Specification of smoking status)
14. study\_complt\_stat (Specification of study completion status, if no the reason should be specified)
15. travel\_out\_six\_month (Specification of the countries travelled in the last six months; can include multiple travels)
16. twin\_sibling (Specification of twin sibling presence)
17. urine\_collect\_meth (Specification of urine collection method)
18. urogenit\_tract\_disor (History of urogenitaltract disorders; can include multiple disorders)
19. weight\_loss\_3\_month (Specification of weight loss in the last three months, if yes should be further specified to include amount of weight loss)

Biosample that is using the MIxS human-associated package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/10102948>

<https://www.ncbi.nlm.nih.gov/biosample/SAMN02808661>

**5. MIxS Human-gut package**

The MIxS Human-gut package provides a list of 127 descriptors to describe the sample from the human gut / digestive system (stool / feces or mucus or tissue or any other sample from the digestive system) used for sequencing.

The **mandatory** descriptors of the human-gut package are the same ten as given in the shared mandatory descriptors in **page 1** of this document.

Further, the following three descriptors are **unique** to this package:

1. gastrointest\_disord (History of gastrointestinal tract disorders; can include multiple disorders )
2. liver\_disord (History of liver disorders; can include multiple disorders)
3. special\_diet (Specification of special diet; can include multiple special diets)

Biosample that is using the MIxS human-gut package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/SAMN02711913>

<https://www.ncbi.nlm.nih.gov/biosample/9258455>

**6. MIxS Human-oral Package**

The MIxS Human-oral package contains 126 parameters to describe the sample that was collected from the human oral cavity including saliva, gingival, dental plaque.

The **mandatory** descriptors of the human-oral package are the same ten as given in the shared mandatory descriptors in **page 1**.

Further, the following two descriptors are **unique** to this package:

1. nose\_mouth\_teeth\_throat\_disord (History of nose/mouth/teeth/throat disorders; can include multiple disorders)
2. time\_last\_toothbrush (Specification of the time since last toothbrushing)

Example Biosamples that are using the MIxS Human-oral package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/2336750>

<https://www.ncbi.nlm.nih.gov/biosample/SAMN02333701>

**7. MIxS Human-skin package**

The MIxS Human-skin package has 127 descriptors for the human skin sample used for sequencing.

The **mandatory** descriptors for human-skin package are the same ten as given in the shared mandatory descriptors in **page 1**.

Further, the following three descriptors are **unique** to the package when compared with other 16 MIxS packages:

1. dermatology\_disord (History of dermatology disorders; can include multiple disorders)
2. dominant\_hand (Dominant hand of the subject)
3. time\_since\_last\_wash (Specification of the time since last wash)

An example biosample that uses the MIxS human-skin can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/SAMN01894483>

**8. The MIxS Human-vaginal package**

The MIxS human-vaginal package contains 134 descriptors to describe the sample collected from the human vagina including vaginal swab, vaginal lavage, mucous membrane, etc.

The **mandatory** descriptors for MIxS human-vaginal package are the same ten as given in the shared mandatory descriptors in **page 1**.

Further, the following 10 descriptors are **unique** to the MIxS human- vaginal package:

1. birth\_control (Specification of birth control medication used)
2. douche (Date of most recent douche)
3. gynecologic\_disord (History of gynecological disorders; can include multiple disorders)
4. hrt (Whether subject had hormone replacement theraphy, and if yes start date)
5. hysterectomy (Specification of whether hysterectomy was performed)
6. menarche (Date of most recent menstruation)
7. menopause (Date of onset of menopause)
8. pregnancy (Date due of pregnancy)
9. sexual\_act (Current sexual partner and frequency of sex)
10. urogenit\_disord (History of urogenital disorders, can include multiple disorders)

Example biosamples that use the MIxS human-vaginal can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/10878090>

<https://www.ncbi.nlm.nih.gov/biosample/14363795>

**9. The MIxS Hydrocarbon resources-cores package**

The MIxS Hydrocarbon resources-cores package contains 173 descriptors to describe the sample related to hydrocarbon resources including oil, gas reservoir and coalbed.

In addition to the ten shared mandatory descriptors givenin **page 1**, another ten descriptors are **mandatory** for the MIxS Hydrocarbon resources-cores package, that are given below:

1. hcr (Main Hydrocarbon Resource type. The term "Hydrocarbon Resource" HCR defined as a natural environmental feature containing large amounts of hydrocarbons at high concentrations potentially suitable for commercial exploitation. This term should not be confused with the Hydrocarbon Occurrence term which also includes hydrocarbon-rich environments with currently limited commercial interest such as seeps, outcrops, gas hydrates etc. If "other" is specified, please propose entry in "additional info" field)
2. hc\_produced (Main hydrocarbon type produced from resource (i.e. Oil, gas, condensate, etc). If "other" is specified, please propose entry in "additional info" field)
3. basin (Name of the basin (e.g. Campos))
4. hcr\_temp (Original temperature of the hydrocarbon resource)
5. sulfate\_fw (Original sulfate concentration in the hydrocarbon resource)
6. vfa\_fw (Original volatile fatty acid concentration in the hydrocarbon resource)
7. samp\_type (Type of material (i.e. sample) collected. Includes types like core, rock trimmings, drill cuttings, piping section, coupon, pigging debris, solid deposit, produced fluid, produced water, injected water, swabs, etc. If "other" is specified, please propose entry in "additional info" field)
8. temp (Temperature of the sample at the time of sampling)
9. samp\_salinity (Salinity is the total concentration of all dissolved salts in a liquid or solid (in the form of an extract obtained by centrifugation) sample. While salinity can be measured by a complete chemical analysis, this method is difficult and time consuming. More often, it is instead derived from the conductivity measurement. This is known as practical salinity. These derivations compare the specific conductance of the sample to a salinity standard such as seawater)
10. api (API gravity is a measure of how heavy or light a petroleum liquid is compared to water (source: https://en.wikipedia.org/wiki/API\_gravity) (e.g. 31.1° API) )

Further, this package contains the following eight **unique** (not present in other 16 MIxS packages) descriptors:

1. owc\_tvdss (Depth of the original oil water contact (OWC) zone (average) (m TVDSS))
2. permeability (Measure of the ability of a hydrocarbon resource to allow fluids to pass through it. (Additional information: https://en.wikipedia.org/wiki/Permeability\_(earth\_sciences)))
3. samp\_md (In non deviated well, measured depth is equal to the true vertical depth, TVD (TVD=TVDSS plus the reference or datum it refers to). In deviated wells, the MD is the length of trajectory of the borehole measured from the same reference or datum. Common datums used are ground level (GL), drilling rig floor (DF), rotary table (RT), kelly bushing (KB) and mean sea level (MSL). If "other" is specified, please propose entry in "additional info" field )
4. samp\_tvdss (Depth of the sample i.e. The vertical distance between the sea level and the sampled position in the subsurface. Depth can be reported as an interval for subsurface samples e.g. 1325.75-1362.25 m)
5. sr\_dep\_env (Source rock depositional environment (https://en.wikipedia.org/wiki/Source\_rock). If "other" is specified, please propose entry in "additional info" field)
6. sr\_geol\_age (Geological age of source rock (Additional info: https://en.wikipedia.org/wiki/Period\_(geology)). If "other" is specified, please propose entry in "additional info" field)
7. sr\_kerog\_type (Origin of kerogen. Type I: Algal (aquatic), Type II: planktonic and soft plant material (aquatic or terrestrial), Type III: terrestrial woody/ fibrous plant material (terrestrial), Type IV: oxidized recycled woody debris (terrestrial) (additional information: https://en.wikipedia.org/wiki/Kerogen). If "other" is specified, please propose entry in "additional info" field)
8. sr\_lithology (Lithology of source rock (https://en.wikipedia.org/wiki/Source\_rock). If "other" is specified, please propose entry in "additional info" field)

Further, when compared specifically with ‘Hydrocarbon resources-**fluids/swabs** package’, the ‘Hydrocarbon resources-**cores** package’ contains ten descriptors as unique (i.e., 163 of the total 173 descriptors of Hydrocarbon resources-**cores** package are also present in the Hydrocarbon resources-**fluids/swabs** package); based on the applicability, these ten descriptors (given below) can be used to assign the ‘MIxS Hydrocarbon resources-**cores** package’ over ‘Hydrocarbon resources-**fluids/swabs** package’ for the studied sample.

1. permeability (Measure of the ability of a hydrocarbon resource to allow fluids to pass through it. (Additional information: https://en.wikipedia.org/wiki/Permeability\_(earth\_sciences)))
2. porosity (Porosity of deposited sediment is volume of voids divided by the total volume of sample)
3. owc\_tvdss (Depth of the original oil water contact (OWC) zone (average) (m TVDSS))
4. sr\_kerog\_type (Origin of kerogen. Type I: Algal (aquatic), Type II: planktonic and soft plant material (aquatic or terrestrial), Type III: terrestrial woody/ fibrous plant material (terrestrial), Type IV: oxidized recycled woody debris (terrestrial) (additional information: https://en.wikipedia.org/wiki/Kerogen). If "other" is specified, please propose entry in "additional info" field)
5. sr\_lithology (Lithology of source rock (https://en.wikipedia.org/wiki/Source\_rock). If "other" is specified, please propose entry in "additional info" field)
6. sr\_dep\_env (Source rock depositional environment (https://en.wikipedia.org/wiki/Source\_rock). If "other" is specified, please propose entry in "additional info" field)
7. sr\_geol\_age (Geological age of source rock (Additional info: https://en.wikipedia.org/wiki/Period\_(geology)). If "other" is specified, please propose entry in "additional info" field)
8. samp\_tvdss (Depth of the sample i.e. The vertical distance between the sea level and the sampled position in the subsurface. Depth can be reported as an interval for subsurface samples e.g. 1325.75-1362.25 m)
9. samp\_md (In non deviated well, measured depth is equal to the true vertical depth, TVD (TVD=TVDSS plus the reference or datum it refers to). In deviated wells, the MD is the length of trajectory of the borehole measured from the same reference or datum. Common datums used are ground level (GL), drilling rig floor (DF), rotary table (RT), kelly bushing (KB) and mean sea level (MSL). If "other" is specified, please propose entry in "additional info" field )
10. elev (Elevation of the sampling site is its height above a fixed reference point, most commonly the mean sea level. Elevation is mainly used when referring to points on the earth's surface, while altitude is used for points above the surface, such as an aircraft in flight or a spacecraft in orbit)

**10. The Hydrocarbon resources-fluids/swabs package**

The Hydrocarbon resources-fluids/swabs package contains 177 descriptors to describe the hydrocarbon fluid or swab sample such as fluids from gas or oil well, drilling fluid, tailing pond.

In addition to the ten shared mandatory descriptors given in **page 1**, 14 more descriptors are **mandatory** for this package (of which 4 are also **unique**):

1. hcr (Main Hydrocarbon Resource type. The term "Hydrocarbon Resource" HCR defined as a natural environmental feature containing large amounts of hydrocarbons at high concentrations potentially suitable for commercial exploitation. This term should not be confused with the Hydrocarbon Occurrence term which also includes hydrocarbon-rich environments with currently limited commercial interest such as seeps, outcrops, gas hydrates etc. If "other" is specified, please propose entry in "additional info" field)
2. hc\_produced (Main hydrocarbon type produced from resource (i.e. Oil, gas, condensate, etc). If "other" is specified, please propose entry in "additional info" field)
3. basin (Name of the basin (e.g. Campos))
4. water\_cut (Current amount of water (%) in a produced fluid stream; or the average of the combined streams) **(unique and mandatory)**
5. iwf (Proportion of the produced fluids derived from injected water at the time of sampling. (e.g. 87%)) **(unique and mandatory)**
6. add\_recov\_method (Additional (i.e. Secondary, tertiary, etc.) recovery methods deployed for increase of hydrocarbon recovery from resource and start date for each one of them. If "other" is specified, please propose entry in "additional info" field) **(unique and mandatory)**
7. samp\_type (Type of material (i.e. sample) collected. Includes types like core, rock trimmings, drill cuttings, piping section, coupon, pigging debris, solid deposit, produced fluid, produced water, injected water, swabs, etc. If "other" is specified, please propose entry in "additional info" field)
8. samp\_collection\_point (Sampling point on the asset were sample was collected (e.g. Wellhead, storage tank, separator, etc). If "other" is specified, please propose entry in "additional info" field) **(unique and mandatory)**
9. temp (Temperature of the sample at the time of sampling)
10. samp\_salinity (Salinity is the total concentration of all dissolved salts in a liquid or solid (in the form of an extract obtained by centrifugation) sample. While salinity can be measured by a complete chemical analysis, this method is difficult and time consuming. More often, it is instead derived from the conductivity measurement. This is known as practical salinity. These derivations compare the specific conductance of the sample to a salinity standard such as seawater)
11. sulfate (Concentration of sulfate in the sample)
12. sulfide (Concentration of sulfide in the sample)
13. nitrate (Concentration of nitrate in the sample)
14. api (API gravity is a measure of how heavy or light a petroleum liquid is compared to water (source: https://en.wikipedia.org/wiki/API\_gravity) (e.g. 31.1° API) )

Further, 14 descriptors are **unique** (not present in other 16 MIxS packages) to this package (of which four are also **mandatory**) and listed below:

1. add\_recov\_method (Additional (i.e. Secondary, tertiary, etc.) recovery methods deployed for increase of hydrocarbon recovery from resource and start date for each one of them. If "other" is specified, please propose entry in "additional info" field) **(unique and mandatory)**
2. biocide (List of biocides (commercial name of product and supplier) and date of administration)
3. biocide\_admin\_method (Method of biocide administration (dose, frequency, duration, time elapsed between last biociding and sampling) (e.g. 150 mg/l; weekly; 4 hr; 3 days))
4. chem\_treatment (List of chemical compounds administered upstream the sampling location where sampling occurred (e.g. Glycols, H2S scavenger, corrosion and scale inhibitors, demulsifiers, and other production chemicals etc.). The commercial name of the product and name of the supplier should be provided. The date of administration should also be included)
5. chem\_treatment\_method (Method of chemical administration(dose, frequency, duration, time elapsed between administration and sampling) (e.g. 50 mg/l; twice a week; 1 hr; 0 days))
6. iw\_bt\_date\_well (Injection water breakthrough date per well following a secondary and/or tertiary recovery)
7. iwf (Proportion of the produced fluids derived from injected water at the time of sampling. (e.g. 87%)) **(unique and mandatory)**
8. prod\_rate (Oil and/or gas production rates per well (e.g. 524 m3 / day))
9. prod\_start\_date (Date of field's first production)
10. samp\_collection\_point (Sampling point on the asset were sample was collected (e.g. Wellhead, storage tank, separator, etc). If "other" is specified, please propose entry in "additional info" field) **(unique and mandatory)**
11. samp\_loc\_corr\_rate (Metal corrosion rate is the speed of metal deterioration due to environmental conditions. As environmental conditions change corrosion rates change accordingly. Therefore, long term corrosion rates are generally more informative than short term rates and for that reason they are preferred during reporting. In the case of suspected MIC, corrosion rate measurements at the time of sampling might provide insights into the involvement of certain microbial community members in MIC as well as potential microbial interplays)
12. samp\_preserv (Preservative added to the sample (e.g. Rnalater, alcohol, formaldehyde, etc.). Where appropriate include volume added (e.g. Rnalater; 2 ml))
13. water\_cut (Current amount of water (%) in a produced fluid stream; or the average of the combined streams) **(unique and mandatory)**
14. water\_production\_rate (Water production rates per well (e.g. 987 m3 / day))

Further, when compared specifically with ‘Hydrocarbon resources-**cores** package’, the ‘Hydrocarbon resources-**fluids/swabs** package’ contains 14 descriptors as unique (i.e., 163 of the total 177 descriptors of Hydrocarbon resources-**fluids/swabs** package are also present in the Hydrocarbon resources-**cores** package); these 14 unique descriptors (given below) can be used to assign the ‘Hydrocarbon resources-**fluids/swabs** package’ over 'Hydrocarbon resources-**cores** package' for the studied sample:

1. prod\_start\_date (Date of field's first production)
2. prod\_rate (Oil and/or gas production rates per well (e.g. 524 m3 / day))
3. water\_production\_rate (Water production rates per well (e.g. 987 m3 / day))
4. water\_cut (Current amount of water (%) in a produced fluid stream; or the average of the combined streams)
5. iwf (Proportion of the produced fluids derived from injected water at the time of sampling. (e.g. 87%))
6. add\_recov\_method (Additional (i.e. Secondary, tertiary, etc.) recovery methods deployed for increase of hydrocarbon recovery from resource and start date for each one of them. If "other" is specified, please propose entry in "additional info" field)
7. iw\_bt\_date\_well (Injection water breakthrough date per well following a secondary and/or tertiary recovery)
8. biocide (List of biocides (commercial name of product and supplier) and date of administration)
9. biocide\_admin\_method (Method of biocide administration (dose, frequency, duration, time elapsed between last biociding and sampling) (e.g. 150 mg/l; weekly; 4 hr; 3 days))
10. chem\_treatment (List of chemical compounds administered upstream the sampling location where sampling occurred (e.g. Glycols, H2S scavenger, corrosion and scale inhibitors, demulsifiers, and other production chemicals etc.). The commercial name of the product and name of the supplier should be provided. The date of administration should also be included)
11. chem\_treatment\_method (Method of chemical administration(dose, frequency, duration, time elapsed between administration and sampling) (e.g. 50 mg/l; twice a week; 1 hr; 0 days))
12. samp\_loc\_corr\_rate (Metal corrosion rate is the speed of metal deterioration due to environmental conditions. As environmental conditions change corrosion rates change accordingly. Therefore, long term corrosion rates are generally more informative than short term rates and for that reason they are preferred during reporting. In the case of suspected MIC, corrosion rate measurements at the time of sampling might provide insights into the involvement of certain microbial community members in MIC as well as potential microbial interplays)
13. samp\_collection\_point (Sampling point on the asset were sample was collected (e.g. Wellhead, storage tank, separator, etc). If "other" is specified, please propose entry in "additional info" field)
14. samp\_preserv (Preservative added to the sample (e.g. Rnalater, alcohol, formaldehyde, etc.). Where appropriate include volume added (e.g. Rnalater; 2 ml))

**11. MIxS Microbial mat/biofilm package**

MIxS Microbial mat/biofilm package has 155 descriptors to describe samples from microbial mat or microbial biofilms such as deep subsurface biofilm, stromatolite mat, glaciers and ice sheets.

Further, this package has no additional mandatory descriptors other than those ten shared mandatory descriptors given in **page 1**.

This package does not have any **unique** descriptors when compared with other 16 packages.

However, when compared with water package, five descriptors are unique in MIxS microbial mat/biofilm package (150 of the total 155 Microbial mat/biofilm descriptors are present in the Water package); these five descriptors present in MIxS microbial mat/biofilm package but not in water package given below might be useful when selecting the MIxS Microbial mat/biofilm package over MIxS water package.

1. methane (Methane (gas) amount or concentration at the time of sampling)
2. tot\_carb (Total carbon content)
3. water\_content (Water content measurement)
4. tot\_org\_carb (Definition for soil: total organic carbon content of the soil, definition otherwise: total organic carbon content)
5. tot\_nitro\_content (Total nitrogen content of the sample)

Example biosample that uses the MIxS Microbial mat/biofilm package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/2740840>

<https://www.ncbi.nlm.nih.gov/biosample/SAMN02213668>

https://www.ncbi.nlm.nih.gov/biosample/14343689

**12. MIxS Miscellaneous natural or artificial environment package (MIxS-misc nat\_artif\_env)**

The Miscellaneous natural or artificial environment package contains a total of 137 descriptors to describe the biosample that is not listed in other packages; for example, cathode or synthetic sample.

Further, this package has no additional **mandatory** descriptors other than those ten shared mandatory descriptors that are provided in the **page 1** of this document.

This package has no **unique** descriptorswhen compared with other 16 MIxS packages.

Biosamples that use the MIxS Miscellaneous natural or artificial environment package package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/13324458>

<https://www.ncbi.nlm.nih.gov/biosample/7550897>

**13. MIxS Plant-associated package**

The MIxS Plant-associated package has 166 descriptors to describe the sample that are from plants [for example: root (root metagenome), leaf (phyllosphere metagenome), rhizosphere (rhizosphere metagenome) or seed (seed metagenome)].

The MIxS Plant-associated package has the same ten shared **mandatory** descriptors that are provided in the **page 1** of this document.

Further, the following 42 descriptors are **unique** to this package:

1. air\_temp\_regm (Information about treatment involving an exposure to varying temperatures; should include the temperature, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include different temperature regimens)
2. antibiotic\_regm (Information about treatment involving antibiotic administration; should include the name of antibiotic, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple antibiotic regimens)
3. biotic\_regm (Information about treatment(s) involving use of biotic factors, such as bacteria, viruses or fungi.)
4. chem\_mutagen (Treatment involving use of mutagens; should include the name of mutagen, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple mutagen regimens)
5. climate\_environment (Treatment involving an exposure to a particular climate; treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple climates)
6. cult\_root\_med (Name or reference for the hydroponic or in vitro culture rooting medium; can be the name of a commonly used medium or reference to a specific medium, e.g. Murashige and Skoog medium. If the medium has not been formally published, use the rooting medium descriptors.)
7. fertilizer\_regm (Information about treatment involving the use of fertilizers; should include the name of fertilizer, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple fertilizer regimens)
8. fungicide\_regm (Information about treatment involving use of fungicides; should include the name of fungicide, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple fungicide regimens)
9. gaseous\_environment (Use of conditions with differing gaseous environments; should include the name of gaseous compound, amount administered, treatment duration, interval and total experimental duration; can include multiple gaseous environment regimens)
10. gravity (Information about treatment involving use of gravity factor to study various types of responses in presence, absence or modified levels of gravity; treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple treatments)
11. growth\_facil (Type of facility where the sampled plant was grown; controlled vocabulary: growth chamber, open top chamber, glasshouse, experimental garden, field. Alternatively use Crop Ontology (CO) terms, see http://www.cropontology.org/ontology/CO\_715/Crop%20Research)
12. growth\_habit (Characteristic shape, appearance or growth form of a plant species)
13. growth\_hormone\_regm (Information about treatment involving use of growth hormones; should include the name of growth hormone, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple growth hormone regimens)
14. herbicide\_regm (Information about treatment involving use of herbicides; information about treatment involving use of growth hormones; should include the name of herbicide, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple regimens)
15. host\_wet\_mass (Measurement of wet mass)
16. humidity\_regm (Information about treatment involving an exposure to varying degree of humidity; information about treatment involving use of growth hormones; should include amount of humidity administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple regimens)
17. light\_regm (Information about treatment(s) involving exposure to light, including both light intensity and quality.)
18. mechanical\_damage (Information about any mechanical damage exerted on the plant; can include multiple damages and sites)
19. mineral\_nutr\_regm (Information about treatment involving the use of mineral supplements; should include the name of mineral nutrient, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple mineral nutrient regimens)
20. non\_mineral\_nutr\_regm (Information about treatment involving the exposure of plant to non-mineral nutrient such as oxygen, hydrogen or carbon; should include the name of non-mineral nutrient, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple non-mineral nutrient regimens)
21. pesticide\_regm (Information about treatment involving use of insecticides; should include the name of pesticide, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple pesticide regimens)
22. ph\_regm (Information about treatment involving exposure of plants to varying levels of ph of the growth media, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple regimen)
23. plant\_growth\_med (Specification of the media for growing the plants or tissue cultured samples, e.g. soil, aeroponic, hydroponic, in vitro solid culture medium, in vitro liquid culture medium. Recommended value is a specific value from EO:plant growth medium (follow this link for terms http://purl.obolibrary.org/obo/EO\_0007147) or other controlled vocabulary)
24. plant\_product (Substance produced by the plant, where the sample was obtained from)
25. plant\_sex (Sex of the reproductive parts on the whole plant, e.g. pistillate, staminate, monoecieous, hermaphrodite.)
26. plant\_struc (Name of plant structure the sample was obtained from; for Plant Ontology (PO) (v releases/2017-12-14) terms, see http://purl.bioontology.org/ontology/PO, e.g. petiole epidermis (PO\_0000051). If an individual flower is sampled, the sex of it can be recorded here.)
27. radiation\_regm (Information about treatment involving exposure of plant or a plant part to a particular radiation regimen; should include the radiation type, amount or intensity administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple radiation regimens)
28. rainfall\_regm (Information about treatment involving an exposure to a given amount of rainfall, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple regimens)
29. root\_cond (Relevant rooting conditions such as field plot size, sowing density, container dimensions, number of plants per container.)
30. root\_med\_carbon (Source of organic carbon in the culture rooting medium; e.g. sucrose.)
31. root\_med\_macronutr (Measurement of the culture rooting medium macronutrients (N,P, K, Ca, Mg, S); e.g. KH2PO4 (170 mg/L).)
32. root\_med\_micronutr (Measurement of the culture rooting medium micronutrients (Fe, Mn, Zn, B, Cu, Mo); e.g. H3BO3 (6.2 mg/L).)
33. root\_med\_ph (pH measurement of the culture rooting medium; e.g. 5.5.)
34. root\_med\_regl (Growth regulators in the culture rooting medium such as cytokinins, auxins, gybberellins, abscisic acid; e.g. 0.5 mg/L NAA.)
35. root\_med\_solid (Specification of the solidifying agent in the culture rooting medium; e.g. agar.)
36. root\_med\_suppl (Organic supplements of the culture rooting medium, such as vitamins, amino acids, organic acids, antibiotics activated charcoal; e.g. nicotinic acid (0.5 mg/L).)
37. salt\_regm (Information about treatment involving use of salts as supplement to liquid and soil growth media; should include the name of salt, amount administered, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple salt regimens)
38. season\_environment (Treatment involving an exposure to a particular season (e.g. Winter, summer, rabi, rainy etc.), treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment)
39. standing\_water\_regm (Treatment involving an exposure to standing water during a plant's life span, types can be flood water or standing water, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple regimens)
40. tiss\_cult\_growth\_med (Description of plant tissue culture growth media used)
41. water\_temp\_regm (Information about treatment involving an exposure to water with varying degree of temperature, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple regimens)
42. watering\_regm (Information about treatment involving an exposure to watering frequencies, treatment regimen including how many times the treatment was repeated, how long each treatment lasted, and the start and end time of the entire treatment; can include multiple regimens)

Biosamples that use the MIxS Plant-associated package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/SAMN02918283>

<https://www.ncbi.nlm.nih.gov/biosample/14254835>

**14. MIxS-Sediment Package**

MIxS Sediment package contains 161 descriptors to describe various sediment samples such as marine sediment, lake sediment, sediment from mangrove biome/ tidal mudflat,

river sediment that were used for sequencing.

In addition to the ten shared mandatory descriptors provided in **page 1**, **“depth”** is also a mandatory descriptor for the Sediment package.

*(****Depth*** *is defined as the vertical distance below local surface. Depth can be reported as an interval for subsurface samples)*

Further, two of the 161 descriptors are **unique** to the sediment package:

1. particle\_class (Particles are classified, based on their size, into six general categories:clay, silt, sand, gravel, cobbles, and boulders; should include amount of particle preceded by the name of the particle type; can include multiple values)
2. sediment\_type (Information about the sediment type based on major constituents)

It should also be noted that when compared to the Water package, only 8 descriptors are unique to the Sediment package (*153 of the 162 Sediment package descriptors are also present in the MIxS Water package*). Therefore, the unique descriptors in the Sediment package may be helpful when choosing the sediment package over water package. Further, when compared with the Soil package, there are 60 unique descriptors present in the Sediment package. If conflict occurs over the package assignment between soil and sediment packages, the unique descriptors available in these packages shall be used for reference.

Biosamples that use the MIxS-Sediment package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/SAMN02892988>

<https://www.ncbi.nlm.nih.gov/biosample/14062185>

**15. MIxS Soil Package:**

MIxS Soil Package contains a list of 147 descriptors to describe the soil sample taken from various environments including soil from, cropland, dryland, forest, grassland soil , coastal sand dune, permafrost soil.

Addition to the shared mandatory descriptors given in **page 1**, **'depth'** and **'elevation'** are mandatory descriptors in the Soil package.

*(****Depth*** *is defined as the vertical distance below local surface. Depth can be reported as an interval for subsurface samples.*

***Elevation*** *of the sampling site is its height above a fixed reference point, most commonly the mean sea level. Elevation is mainly used when referring to points on the earth's surface.)*

Further, the following 46 descriptors are **unique** in the soil package when compared with other MIxS packages:

1. agrochem\_addition (Addition of fertilizers, pesticides, etc. - amount and time of applications)
2. al\_sat (Aluminum saturation (esp. For tropical soils))
3. al\_sat\_meth (Reference or method used in determining Al saturation)
4. annual\_precpt (The average of all annual precipitation values known, or an estimated equivalent value derived by such methods as regional indexes or Isohyetal maps. )
5. annual\_temp (Mean annual temperature)
6. crop\_rotation (Whether or not crop is rotated, and if yes, rotation schedule)
7. cur\_land\_use (Present state of sample site)
8. cur\_vegetation (Vegetation classification from one or more standard classification systems, or agricultural crop)
9. cur\_vegetation\_meth (Reference or method used in vegetation classification )
10. drainage\_class (Drainage classification from a standard system such as the USDA system)
11. extreme\_event (Unusual physical events that may have affected microbial populations)
12. extreme\_salinity (Measured salinity )
13. fao\_class (Soil classification from the FAO World Reference Database for Soil Resources. The list can be found at http://www.fao.org/nr/land/sols/soil/wrb-soil-maps/reference-groups)
14. fire (Historical and/or physical evidence of fire)
15. flooding (Historical and/or physical evidence of flooding)
16. heavy\_metals (Heavy metals present and concentrationsany drug used by subject and the frequency of usage; can include multiple heavy metals and concentrations)
17. heavy\_metals\_meth (Reference or method used in determining heavy metals)
18. horizon (Specific layer in the land area which measures parallel to the soil surface and possesses physical characteristics which differ from the layers above and beneath)
19. horizon\_meth (Reference or method used in determining the horizon)
20. link\_addit\_analys (Link to additional analysis results performed on the sample)
21. link\_class\_info (Link to digitized soil maps or other soil classification information)
22. link\_climate\_info (Link to climate resource)
23. local\_class (Soil classification based on local soil classification system)
24. local\_class\_meth (Reference or method used in determining the local soil classification )
25. microbial\_biomass (The part of the organic matter in the soil that constitutes living microorganisms smaller than 5-10 micrometer. If you keep this, you would need to have correction factors used for conversion to the final units)
26. microbial\_biomass\_meth (Reference or method used in determining microbial biomass)
27. ph\_meth (Reference or method used in determining ph)
28. pool\_dna\_extracts (Indicate whether multiple DNA extractions were mixed. If the answer yes, the number of extracts that were pooled should be given)
29. previous\_land\_use (Previous land use and dates)
30. previous\_land\_use\_meth (Reference or method used in determining previous land use and dates)
31. profile\_position (Cross-sectional position in the hillslope where sample was collected.sample area position in relation to surrounding areas)
32. salinity\_meth (Reference or method used in determining salinity)
33. season\_precpt (The average of all seasonal precipitation values known, or an estimated equivalent value derived by such methods as regional indexes or Isohyetal maps. )
34. season\_temp (Mean seasonal temperature)
35. sieving (Collection design of pooled samples and/or sieve size and amount of sample sieved )
36. slope\_aspect (The direction a slope faces. While looking down a slope use a compass to record the direction you are facing (direction or degrees); e.g., nw or 315 degrees. This measure provides an indication of sun and wind exposure that will influence soil temperature and evapotranspiration.)
37. slope\_gradient (Commonly called 'slope'. The angle between ground surface and a horizontal line (in percent). This is the direction that overland water would flow. This measure is usually taken with a hand level meter or clinometer)
38. soil\_type (Soil series name or other lower-level classification)
39. soil\_type\_meth (Reference or method used in determining soil series name or other lower-level classification)
40. store\_cond (Explain how and for how long the soil sample was stored before DNA extraction)
41. texture (The relative proportion of different grain sizes of mineral particles in a soil, as described using a standard system; express as % sand (50 um to 2 mm), silt (2 um to 50 um), and clay (<2 um) with textural name (e.g., silty clay loam) optional.)
42. texture\_meth (Reference or method used in determining soil texture)
43. tillage (Note method(s) used for tilling)
44. tot\_nitro\_content\_meth (Reference or method used in determining the total nitrogen)
45. tot\_org\_c\_meth (Reference or method used in determining total organic carbon)
46. water\_content\_soil\_meth (Reference or method used in determining the water content of soil)

An example biosamples that use the MIxS-Soil package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/SAMN02899785>

**16. MIxS Wastewater/sludge**

MIxS Wastewater/sludge contains a list of 131 descriptors to describe wastewater / sludge sample including sewage, bioreactor sludge, plumbing drain, municipal wastewater, wastewater treatment plant.

The mandatory descriptors of the MIxS Wastewater/sludge are the same ten as given in the shared **mandatory** descriptors in **page 1** of this document.

Further, there are 19 descriptors are **unique** to this package and given below:

1. biochem\_oxygen\_dem (Amount of dissolved oxygen needed by aerobic biological organisms in a body of water to break down organic material present in a given water sample at certain temperature over a specific time period)
2. chem\_oxygen\_dem (A measure of the capacity of water to consume oxygen during the decomposition of organic matter and the oxidation of inorganic chemicals such as ammonia and nitrite)
3. efficiency\_percent (Percentage of volatile solids removed from the anaerobic digestor)
4. emulsions (Amount or concentration of substances such as paints, adhesives, mayonnaise, hair colorants, emulsified oils, etc.; can include multiple emulsion types)
5. gaseous\_substances (Amount or concentration of substances such as hydrogen sulfide, carbon dioxide, methane, etc.; can include multiple substances)
6. indust\_eff\_percent (Percentage of industrial effluents received by wastewater treatment plant)
7. inorg\_particles (Concentration of particles such as sand, grit, metal particles, ceramics, etc.; can include multiple particles)
8. org\_particles (Concentration of particles such as faeces, hairs, food, vomit, paper fibers, plant material, humus, etc.)
9. pre\_treatment (The process of pre-treatment removes materials that can be easily collected from the raw wastewater)
10. primary\_treatment (The process to produce both a generally homogeneous liquid capable of being treated biologically and a sludge that can be separately treated or processed)
11. reactor\_type (Anaerobic digesters can be designed and engineered to operate using a number of different process configurations, as batch or continuous, mesophilic, high solid or low solid, and single stage or multistage)
12. secondary\_treatment (The process for substantially degrading the biological content of the sewage )
13. sewage\_type (Type of wastewater treatment plant as municipial or industrial)
14. sludge\_retent\_time (The time activated sludge remains in reactor)
15. soluble\_inorg\_mat (Concentration of substances such as ammonia, road-salt, sea-salt, cyanide, hydrogen sulfide, thiocyanates, thiosulfates, etc.)
16. soluble\_org\_mat (Concentration of substances such as urea, fruit sugars, soluble proteins, drugs, pharmaceuticals, etc.)
17. tertiary\_treatment (The process providing a final treatment stage to raise the effluent quality before it is discharged to the receiving environment)
18. tot\_phosphate (Total amount or concentration of phosphate)
19. wastewater\_type (The origin of wastewater such as human waste, rainfall, storm drains, etc.)

Example biosamples that use the MIxS- Wastewater/sludge package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/SAMN02911905>

<https://www.ncbi.nlm.nih.gov/biosample/13222807>

**17. MIxS Water Package**

MIxS water package contains 176 descriptors to describe water biosample that comes from various resources including water from marine, lake, river or groundwater.

Addition to the ten shared mandatory descriptors given in the **page 1** of this document, "**depth**" is also a mandatory descriptor for the MIxS water package.

*(****Depth*** *is defined as the vertical distance below local surface. Depth can be reported as an interval for subsurface samples)*

There are 18 descriptors **unique** to the water package:

1. atmospheric\_data (Measurement of atmospheric data; can include multiple data)
2. bac\_prod (Bacterial production in the water column measured by isotope uptake)
3. bac\_resp (Measurement of bacterial respiration in the water column)
4. conduc (Electrical conductivity of water)
5. diss\_inorg\_nitro (Concentration of dissolved inorganic nitrogen )
6. down\_par (Visible waveband radiance and irradiance measurements in the water column)
7. fluor (Raw or converted fluorescence of water)
8. light\_intensity (Measurement of light intensity)
9. part\_org\_nitro (Concentration of particulate organic nitrogen)
10. photon\_flux (Measurement of photon flux)
11. primary\_prod (Measurement of primary production, generally measured as isotope uptake)
12. size\_frac\_low (Refers to the mesh/pore size used to pre-filter/pre-sort the sample. Materials larger than the size threshold are excluded from the sample)
13. size\_frac\_up (Refers to the mesh/pore size used to retain the sample. Materials smaller than the size threshold are excluded from the sample)
14. soluble\_react\_phosp (Concentration of soluble reactive phosphorus)
15. suspend\_part\_matter (Concentration of suspended particulate matter)
16. tot\_diss\_nitro (Total dissolved nitrogen concentration, reported as nitrogen, measured by: total dissolved nitrogen = NH4 + NO3NO2 + dissolved organic nitrogen)
17. tot\_inorg\_nitro (Total inorganic nitrogen content)
18. tot\_part\_carb (Total particulate carbon content)

An example biosamples that use the MIxS- Water package can be accessed from here:

<https://www.ncbi.nlm.nih.gov/biosample/SAMN00001362>