**Changelog:**

|  |  |  |
| --- | --- | --- |
| ***Date*** | ***Author*** | ***Description*** |
| 2011-09-21 | PH | Initial version of the document |
| 2011-09-21 | MK | Corrections regarding structure details, adding a dictionary, unification of Polish naming, corrections of structure content and inclusion of ROM |
| 2011-11-10 | PH | Edits and corrections |
| 2011-12-19 | PH | The Structure of Medical Data and Open Issues |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Naming and structure of motion data entered into BDR

## Correct file sets and their naming

All files intended to be inserted into BDR directly, i.e. not packaged into a .zip file, must have names that match the regular expression:

\d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}(-T\d{2})? (\.\d+)?\. (asf|amc|c3d|avi|zip|mp|vsk)

It follows that the name of each file should have a two- or three-character extension with one of the names specified in the above-mentioned expression. The name pattern can be distinguished by:

* ID of the motion resource (session or measurement sample) – corresponding to the string \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}(-T\d{2})?
* an optional fragment of the name consisting of a period and a non-empty string of digits (used in .avi files)
* extension

The following additional rules should be satisfied by the motion resource ID:

* \d{4}-\d{2}-\d{2} must represent a valid date in ISO 8601 format and match the actual recording date of the session
* The string B\d{4} must contain, followed by an appropriate number of zeros, a number identifying the patient/person/**performer**/*subject*/***body***. This means that all sessions in which a given performer appears should have an identical B\d{4} segment (if possible – also if they are created as part of separate projects). These numbers are globally unique.
* The string S\d{2} must contain the **session**  *number,* followed by a zero if necessary. These numbers are not globally unique among all sessions, but only in combination with the "B" segment. I.e. there can be any number of sessions no. 1 in the database, but for performer no. 2 there can be no more than one session no. 1 (even if they differ in date). It is advisable that the session numbers for a given performer be assigned in chronological order.
* The string T\d{2} must contain **the** *trial number*  of the session, started with a zero if necessary. Sample numbers are unique within the numbering space defined by the B\d{4}-S\d sequence{2}

The other naming rules for files are as follows:

* The middle segment of the name containing a period and a decimal number is only acceptable for .avi files
* The extensions zip, asf, mp, vsk are intended to appear only in filenames belonging to the session, i.e. without the -T\d segment{2}
* the extensions c3d, amc, avi are intended to appear only in the names of files belonging to the measurement test, i.e. having the -T\d{2} segment or ROM (the latter – being an acronym for *Range Of* Motion – reserved for the calibration test)

An additional requirement is that the contents of the files (including the contents and names of the files inside the .zip file) do not contain personal data of the performers (name or surname).

## Filesets used in BDR at the level of a Measurement Session for a single performer

### Multimodal data

One file of a given pattern per session:

* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}.asf
* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}.mp
* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}.vsk
* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}.zip

One file of a given pattern for each sample:

* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-T\d{2}\.amc
* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-T\d{2}\.c3d

Four files of a given pattern for each sample:

* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-T\d{2}\.\d+\.avi

### Single-modal data (MOCAP)

One file of a given pattern per session:

* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}.asf
* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}.mp
* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-ROM.vsk[[1]](#footnote-1)
* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-ROM.c3d [[2]](#footnote-2)
* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}.zip

One file of a given pattern for each sample:

* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-T\d{2}\.amc
* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-T\d{2}\.c3d

### Single-modal data (EMG, GRF)

One file of a given pattern per session:

* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}.zip

One file of a given pattern for each sample:

* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-T\d{2}\.c3d

### Unimodal data (VIDEO)

One file of a given pattern per session:

* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}.zip

Four files of a given pattern for each sample:

* \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-T\d{2}\.\d+\.avi

## Filesets used in BDR at the Measurement Session level for multiple performers

If there are multiple performers, we tag the session directory with an identifier that stores information about the IDs of all performers who are in the measurement session at the same time. These are identifiers with numbers from B5000.

The sets of files used in BDR at the level of a Measurement Session for multiple performers are analogous to the duplication of the set of files used in BDR at the level of a Measurement Session for a single performer.

## Data created automatically by the Create Session Wizard operation from files

The wizard service – consisting strictly speaking of the auxiliary validation operation available as part of the BasicQueriesWS service and the main wizard operation available as part of FileStoremanWS – provides coarse validation and automatic setting of some of the attributes of uploaded motion resources. Since the wizard service does not parse any of the uploaded files, the wizard will automatically fill in only the values of those attributes that can be determined by analyzing the list of names of the uploaded session files. This means that the following values will be determined automatically:

* A new instance in the table containing sessions
* session name – in the form \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}
* Session date – in the form \d{4}-\d{2}-\d{2}
* performer identifier – as a number extracted from segment B\d{4} – if necessary, an appropriate row is created in the database in the performers table
* associating the created session with the executor identified by the above-mentioned number – an instance of the Executor Configuration connecting the session and the executor is created
* a new instance for each of the attempts of a given session – i.e. for each unique segment of T\d{2} identified in the list of files of a given session
* the name of each attempt – in the form \d{4}-\d{2}-\d{2}-B\d{4}-S\d{2}-T\d{2}
* name all successfully entered files and assign these files as subordinate instances of the Session entity (not containing the "T" segment) or Trial (containing the "T" segment), respectively
* if there are .asf and .amc files among the uploaded files – for each of the .amc files – the value of the SkeletonFile attribute is filled with the identifier of the .asf file (assigned after uploading it to the database)

## Manually filled in data

Other data that is provided by the current schema and remains unfilled or filled with the default value includes:

* Description – for the measurement test
* Tags – for sessions
* Description – for sessions
* Assigning Sessions to Survey Configurations
* A set of anthropometric data – implemented in the general part of the BDR as a group of attributes for the Session entity – their entry is facilitated by the stored procedure feed\_anthropometric\_data.In the case of providing data to be uploaded in the form of files, the agreed set of atropometric data:
  + BodyMass
  + Height
  + InterAsisDistance
  + LeftLegLength
  + RightLegLenght
  + LeftKneeWidth
  + RightKneeWidth
  + LeftAnkleWidth
  + RightAnkleWidth
  + LeftCircuitThigh
  + RightCircuitThight
  + LeftCircuitShank
  + RightCircuitShank
  + LeftShoulderOffset
  + RightShoulderOffset
  + LeftElbowWidth
  + RightElbowWidth
  + LeftWristWidth
  + RightWristWidth
  + LeftWristThickness
  + RightWristThickness
  + LeftHandWidth
  + RightHandWidth
  + LeftHandThickness
  + RightHandThickness

It should be provided in the form of a CSV file, along with identifiers that allow you to associate individual rows with individual sessions.

* assigning a session to a session group (see below)

## How to use Session Groups

A session group is an entity that currently has a name as its only attribute, allowing a many-to-many relationship to be assigned to an instance of the Session entity.

## Medical data

The medical part (M) of the BDR is a stand-alone database that can be used to identify individuals in the anonymized central part of the BDR (cBDR).

At the moment, there are no facilities for entering medical part data through the user interface, nor through web service operations. A complete description of the patient and the examination requires the following steps:

* appoint (if it has not already been introduced during the first examination of a given person) a new instance of the Patient entity (IdPatient, First Name, Last Name, Data\_urodzenia, Gender) – the identifier should correspond to the value assigned in the cBDR to the Performer identifier corresponding to a given patient (i.e. the number in segment B\d{4} of the naming scheme described above),
* associating a patient with any number of Disease Entities (the Pacjent\_Jednostka\_chorobowa entity has the following attributes: IdPatient, IdDisease, Comment, Main[[3]](#footnote-3)
* uploading, as a BLOB attribute, a photo of the patient in the .jpg file,
* create an instance of the Research entity and associate it with a patient; attributes to be filled in include: Date, Description, Notes, Session ID (indication of the session related to the study in cBDR), IdTyp\_badania – indication of one of the defined types of surveys,
* assignment of the introduced Study to at most one Grupy\_badań. Study groups are set up to group together studies with the same authorisation requirements for access to them (e.g. from the same hospital or department)
* assigning the entered study to any number of Kontekstów\_badania. Contexts are defined in terms of the motive/purpose of the study (e.g. treatment of hip degeneration, treatment of spinal degeneration, treatment of Parkinson's disease, etc.). Assigning a study to a context can be provided with the value of the Role attribute, which specifies the role of the test in the treatment process (the value of the attribute can be, for example, "before the procedure")

## Open questions about medical data

Regardless of the need to build facilities to facilitate the entry of the above-mentioned medical data, it is necessary to (a) resolve the methods of grouping and classification and (b) ensure unambiguous interpretation of the data provided (e.g. records in the patient's .enf files), enabling the extraction and proper interpretation of the data in relation to the scheme described above.

### Ways of grouping and classifying

The following issues are to be resolved in the current instance of the medical database:

* what Grupy\_badań should be appointed,
* what Konteksty\_badań will be used,
* What roles of research within each context will be useful.
* Whether and how to supplement the current Jednostek\_chorobowych catalogue, which now includes:
  + Hypertension
  + Spinal degeneration - cervical spine
  + Spinal degeneration - lumbar spine
  + Degeneration of the spine - the entire length
  + Hip joint endoprosthesis - left side
  + Hip joint endoprosthesis - right side
  + Hip qualified for surgery - right
  + Hip qualified for surgery - left
  + Pain in the upper limbs
  + Knee osteoarthritis - right
  + Osteoarthritis of the knee joint - left
  + Right hip osteoarthritis
  + Hip osteoarthritis - left
  + Stroke - right side
  + Stroke - Left Side
  + Damaged meniscus - right
  + Damaged meniscus - left
  + Osteoarthritis
  + Degeneration of the spine
  + Diabetes
  + Knee qualified for endoprosthesis - right
  + Knee qualified for endoprosthesis - left
  + Discopathy
  + Osteoporosis
* Is the binary division of the disease entities identified for a given patient into the main ones from the point of view of therapy and the rest adequate?
* If so, how exactly is this division to be made?

### Ensuring unambiguous interpretation of data

Currently, medical data is provided as a collection of .enf files and images, one for each patient. From the .enf file, it is possible to recover personal data and information about disease entities. Information about the tests is available implicitly: the assumption that the date of the examination is equal to the date included in the name of the corresponding session, the context of the study – results from the disease entities identified for the patient, and the role of the study in a given context – results from the adopted convention (e.g. the first two delivered sessions of a patient with hip osteoarthritis – before the procedure, and the next two – after the procedure). Therefore, this data is much more modest than the descriptive capabilities of the prepared database schema.

Assuming that the resolution of the issues in the previous subsection will not result in changes to the current scheme, the data provided for uploading to the database should unambiguously resolve the following details:

* Which diseases of the patient indicated in the .enf file are to be saved as the main ones.
* (if needed) Where to look for information about the Dacie\_diagnozy and the Commentary concerning a given disease entity diagnosed for the patient.
* To which Profili\_badania and in what role should the introduced test be assigned (it is not always unambiguous – e.g. some patients have been diagnosed with both spinal degeneration and other degenerations).
* (if needed) Where to look for data constituting the Description and Notes for a given study.
* How will it be determined to which Grupy\_badań a given study is to be assigned.

## Dictionary

**Contractor** – a person subjected to measurements in a laboratory

**Measurement Test** – measurement of the contractor's movement

**Measurement session**  – a set of measurement tests

**Calibration Test** – a measurement test during which we calibrate the measuring system

**Unimodal measurement** – measurement made with the help of a single measurement system. The following systems are available in the laboratory: MOCAP, EMG, GRF, VIDEO

**Multimodal measurement** – measurement performed with the use of multiple measurement systems at the same time

**Single-modal data** – data from measurements of a unimodal system, usually performed for the purposes of computer animation. In the case of unimodal data, it is necessary to expand what data is meant: MOCAP, VIDEO, GRF, EMG

**Multimodal data** – data from measurements of a multimodal system, usually performed for medical purposes

1. Only appears in Blade data [↑](#footnote-ref-1)
2. Only appears in Blade data [↑](#footnote-ref-2)
3. This attribute indicates whether a given disease entity is the main motive for examining a given patient in terms of movement analysis. [↑](#footnote-ref-3)