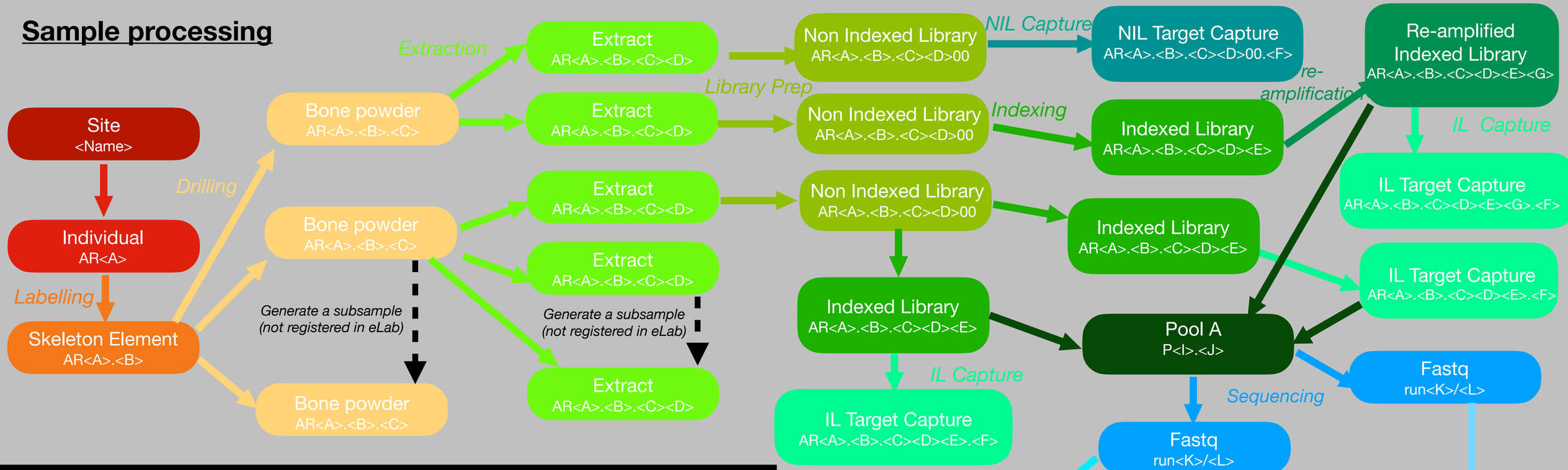
Project Group. Ancient DNA

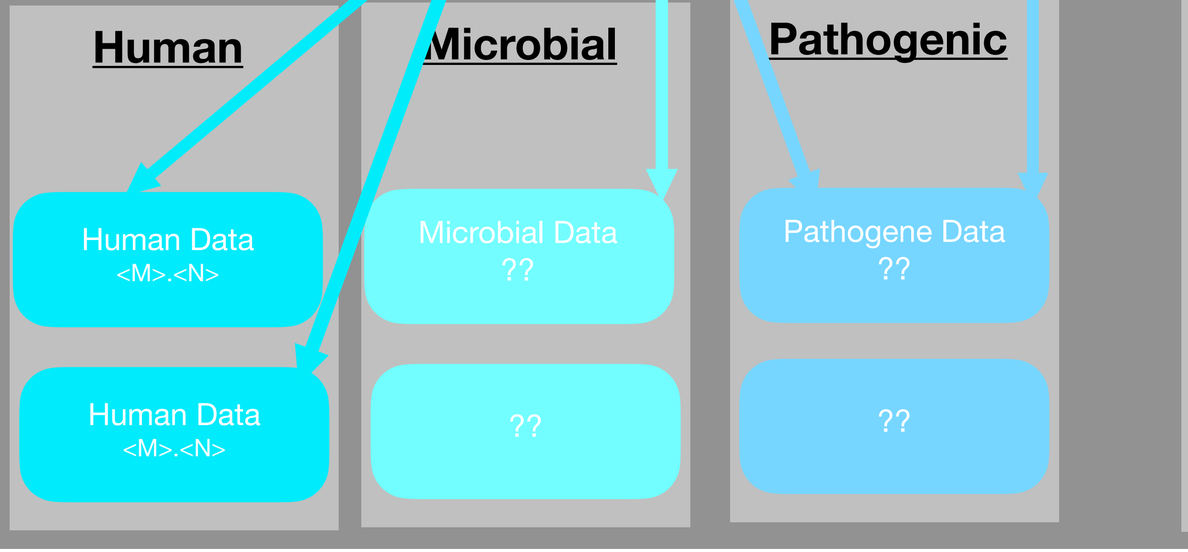
Sample processing



Grey Squares: "Project"
- A set of "Sample types", "Studies" shared for a given goal

Oval: "Sample type"
- One colour per sample type
- IDs with a pattern (explanation next slide).
- AVOID duplicated IDs
- Each sample type have configured fields generic enough so that all information required for a sample (anyhow generated) can be saved
  - Detailed in following slides (or to be defined)

Arrow: "Study" (<=> processing step)
- One "Study" may have as many different "Experiments" (<=> protocoles) as needed:
- e.g. extract can be generated through a "long" or a "short" protocoles. So "Study" Extraction will have two experiments: "Short Incubation" and "Long incubation"



ID system

Individual
AR<A>

<A>: A four digits number. AR can be changed to CH (for chilean samples)

Skeleton Element
AR<A>.<B>

<B>: A one digit number. If more than 10 skeleton element per individual, use letter

Bone powder
AR<A>.<B>.<C>

<C>: A one digit number. If more than 10 bone powder per skeleton element, use letter

Extract
AR<A>.<B>.<C><D>

<D>: A one digit number. If more than 10 extracts per bone powder, use letter

Non Indexed Library
AR<A>.<B>.<C><D>00

Always 00 for Non Indexed Library

<F>: a code designing the kind of target
(.e.g tb, hs, yp, etc…)

NIL Target Capture
AR<A>.<B>.<C><D>00.<F>

Indexed Library
AR<A>.<B>.<C><D><E>

<E>: a two digits number.
Note: Deal with 10 digits restriction for libraries in pool at CGAT,
the combination <D><E> is converted by Alba into a 2 letter code,
generating a new ID for that library (kept in eLab in a field, e.g. ID for pool).
AR0001.1.0102 <-> AR0001.ZC AR0013.1.0102 <-> AR0013.FA

<F>: a code designing the kind of target
(.e.g tb, hs, yp, etc…)
Note: if target on a reamplified library then ID will be
AR<A>.<B>.<C><D><E><G>.<F>

IL Target Capture
AR<A>.<B>.<C><D><E>.<F>

Re-amplified Indexed Library
AR<A>.<B>.<C><D><E><G>

<G>: a letter
Note: Deal with 10 digits restriction for libraries in pool at CGAT,
the combination <D><E> is converted by Alba into a 2 letter code,
generating a new ID for that library (kept in eLab in a field, e.g. ID for pool).
AR0001.1.0102a <-> AR0001.ZCa AR0013.1.0102b <-> AR0013.FAb

Pool A
P<I>.<J>

<I>: Date with format YYMMDD
<J>: One letter

Fastq
run<J>/<L>

<J>: a four digits number (run<J> corresponding to the subfolder in 00_rawdata of the cluster)
<L>: the full name of the files in run<J>. Can content:
- slashes (to mirror the subfolders).
- {1,2} characters if Paired-end sequencing

e.g. run0004/X201SC22040132-Z01-F001_2/AR0014_1/AR0014_1_EKDL220004404-1a-D708-AK1545_HL2J5DSX3_L4_{1,2}.fq.gz

<M>: a string corresponding to the subfolder in Hotpaleo
- slashes (to mirror the subfiolders).
<N>: Date with format YYYY-MM-DD
<O>: A string for an explicit name
e.g. pierre/Projects/2022-03-25_96TartuPetrous_1stScreening

Human Data
<M>/<N>_<O>

Microbial Data
??

Pathogen Data
??