

# GBIS 3.0

## Bakgrund

Det finns en standard GBIS (Gemensam Biobank Informations Standard) 2.0 framtagen sedan tidigare. Tyvärr används denna standarden ej (eller endast i mycket begränsad utsträckning) i praktiken. Detta kan ha flera skäl, men ett av dem torde vara att den ger ganska lite praktisk vägledning i hur den ska appliceras. I arbetet med GBIS 3.0 har jag tagit avstamp i den GBIS 2.0 och försökt att i så stor utsträckning som möjligt återanvända termer och begrepp. Skillnaden i version 3.0 är att den syftar mer till att också beskriva struktur (t.ex. filformat) och inte bara innehåll. Målet är att göra en praktiskt användbar standard som snabbt kan tas i bruk snarare än en komplett nomenklatur för biobanksprovers hela livscykel. Detta innebär att standarden på intet sätt är heltäckande utan är tänkt att utvecklas i nya versioner allteftersom den används, dock med en stabil stomme.

I mitt arbete har jag också tittat på de standarder som arbetas på på andra håll (t.ex. HL7FHIR - specimen) och konstaterar till att de syftar till att lösa problem på en högre nivå, t.ex. annoteringar av resultat och utbyte av stora datamängder snarare än att som här primärt fokusera på ett fysiskt prov, eller är så komplexa att det i realiteten inte kan användas i den praktiska verksamheten i nuläget. Jag tror dock att det finns ett värde för Biobank Sverige att titta på dessa arbeten för framtida standarder. Även den nationella tjänsteplattformen kan bli aktuell att använda, men jag tror att detta arbete ska ses som ett steg på vägen mot att underlätta informationsöverföring både inom och mellan biobanker. Att påbörja ett arbete med att utreda hur biobankerna i framtiden kan ansluta sig till tjänsteplattformen tror jag dock är värdefullt att påbörja, men det ligger långt utanför omfattningen av detta projekt.

## Struktur

GBIS 3.0 utgår från ett fysisk prov och det är också det som är den centrala, betydelsebärande, komponenten.

Man kan naturligtvis representera en standard på olika sätt och jag väljer här på att fokusera på JSON (JavaScript Object Notation, se <https://www.json.org/json-en.html>) när det kommer till praktiska exempel då detta är den absolut vanligaste och mest använda tekniken för informationsöverföring för denna typ av data mellan it-system idag. CSV-filer (se <https://tools.ietf.org/html/rfc4180>) är idag tyvärr fortfarande vanligt förekommande inom den medicinska världen och även om jag tror att det har en enorm nytta att röra sig mot modernare tekniker kommer jag att visa exempel på hur standarden kan se ut om man väljer att arbeta med filer.

Ett prov har följande information:

Fältnamn	Datatyp	Obligatoriskt	Beskrivning
<i>sample_id</i>	Sträng av minst 1, max 30 tecken från ASCII 33 - 126	Ja	Motsvarar den fysiska märkning som unikt identifierar ett rör inom en provsamling
<i>study_id</i>	Sträng av minst 1, max 30 tecken från ASCII 33 - 126	Nej	Identifierar den provsamling vilken provet tillhör. Används (bland annat) för att säkerställa unika identifierare i kombination med <i>sample_id</i>
<i>sample_taken_date</i>	Datum enligt ISO 8601 på formatet YYYY-MM-DD utan tidszonsangivelse	Nej	Datum då provet är taget
<i>sample_usage</i>	Någon av textsträngarna [RE-SEARCH, CARE, RESARCH_AND_CARE]	Ja	Beskriver hur provet får användas
<i>last_usage_date</i>	Datum enligt ISO 8601 på formatet YYYY-MM-DD utan tidszonsangivelse	Nej	Sista dag då provet får användas, t.ex. om dett finns ett etikprövningsbeslut som begränsar i tid.

Fältnamn	Datatyp	Obligatoriskt	Beskrivning
<i>donor</i>	Sträng av minst 1, max 30 tecken från ASCII 33 - 126	Nej	Identifierar provgivaren med t.ex. personnummer eller studienummer. Om personnummer används ska det vara på formatet ÅÅÅÅMMDD-NNNN och ingen hänsyn tas till att personen fyllt hundra år utan skiljetecknet '-' används oavsett.
<i>parent_sample_id</i>	Samma format som <i>sample_id</i>	Nej	Motsvarar den fysiska märkning som unikt identifierar det rör inom en provsamling som detta rör härstammar från
<i>material_type</i>	Sträng av 3 tecken, se separat tabell	Nej	
<i>sample_volume</i>	Heltal mellan 0 och 1000000	Nej	Provmaterialets volym i mikroliter
<i>freeze_thaw_cycles</i>	Heltal mellan 0 och 255	Nej	Beskriver hur många gånger provet tinats och återinfrysas
<i>sample_plate_id</i>	Sträng av minst 1, max 30 tecken från ASCII 33 - 126	Nej	Motsvarar den fysiska märkning som unikt identifierar eventuell platta i vilket provet återfinns

Fältnamn	Datatyp	Obligatoriskt	Beskrivning
<i>sample_plate_position</i>	Sträng av minst 1, max 30 tecken från ASCII 33 - 126	Nej	Provets position i fysisk platta, t.ex. A01

### ***sample\_type***

Sample type ska vara någon av följande

Värde	Beskrivning
BLD	Whole blood
BUF	Unficolled cryopreserved buffy coat, viable
BFF	Unficolled buffy coat, non-viable
CDN	ComplementaryDNA (cDNA)
CEL	PBMC cells, viable
CRD	Cord blood
DWB	Dried whole blood
PEL	PBMC cells, non-viable
PL*	Plasma, all
PL1	Plasma, single-spun
PL2	Plasma, double-spun
RBC	Red blood cells
SER	Serum
XDA	Extracted DNA
WDA	Whole Genome Amplified DNA

Exempel på prov (JSON):

```
{
  sample_id: 12345
  study_id: "HEART_STUDY_2"
  sample_taken_date": "2001-01-01",
  sample_usage: "RESEARCH"
  last_usage_date: "2023-06-12"
  donor: "19121212-1212"
  parent_sample_id: "43425-Y2"
  material_type: "SER"
  sample_volume: 125
  freeze_thaw_cycles: 4
  sample_plate_id: "3245-53"
  sample_plate_position: "A01"
}
```

Exempel på prov (CSV):

```
"sample_id","study_id","sample_taken_date":,"sample_usage","last_usage_date","donor","parent"
"123456","HEART_STUDY_2","2001-01-01","RESEARCH","2023-06-12","19121212-1212","43425-Y2","SP"
```

OM csv-formatet används ska den ha exakt en header-rad enligt ovan och använda teckentabell UTF-8 med CRLF som radavslut. Samtliga fält ska vara quotade med " (även siffror). Escape-character ska vara \ .

Naturligtvis kan en csv-fil innehålla mer än ett prov på samma sätt som en JSON-struktur. OM JSON-strukturen sparas till fil ska denna fil använda teckentabell UTF-8 och radavslut CRLF.