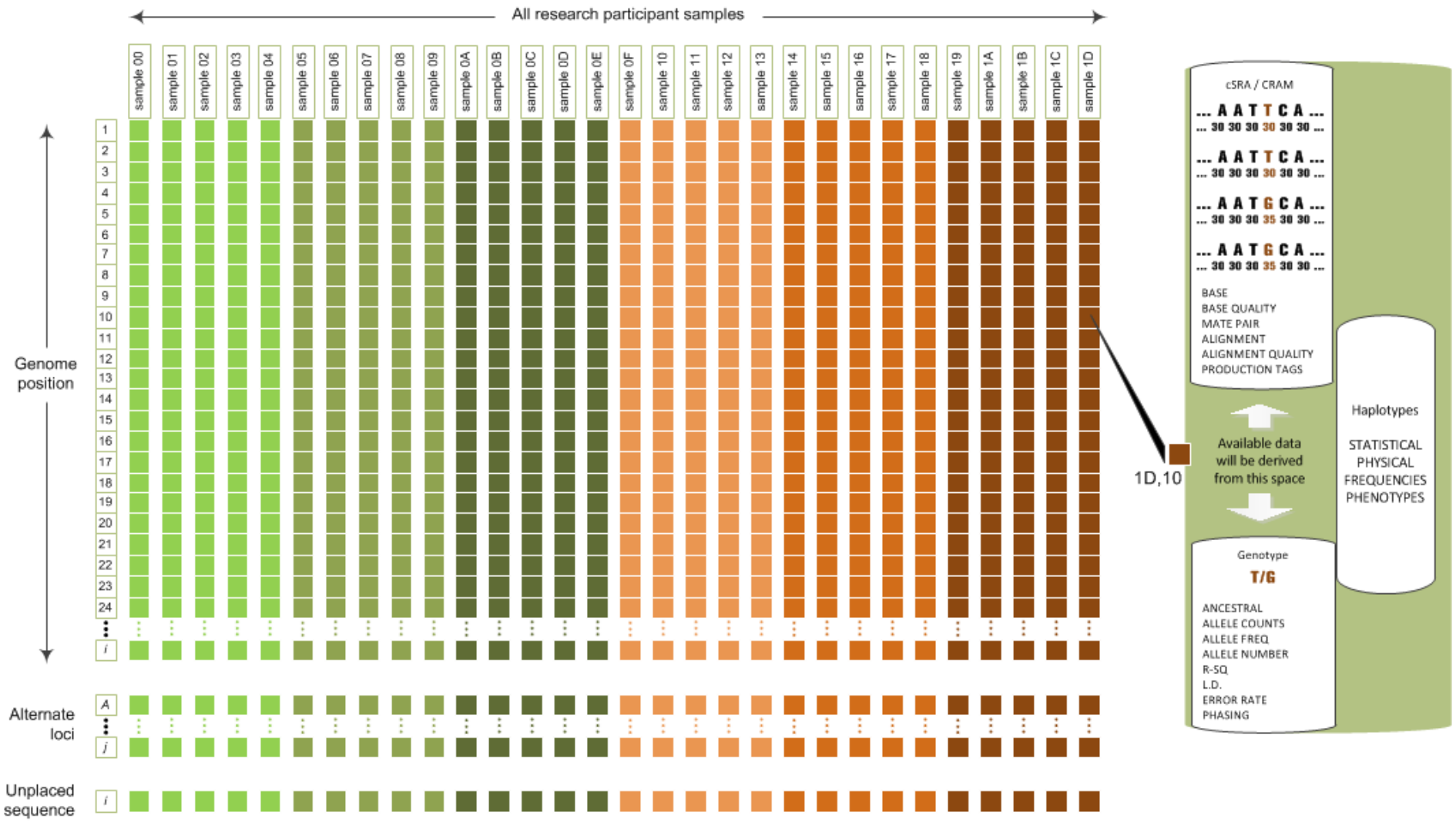
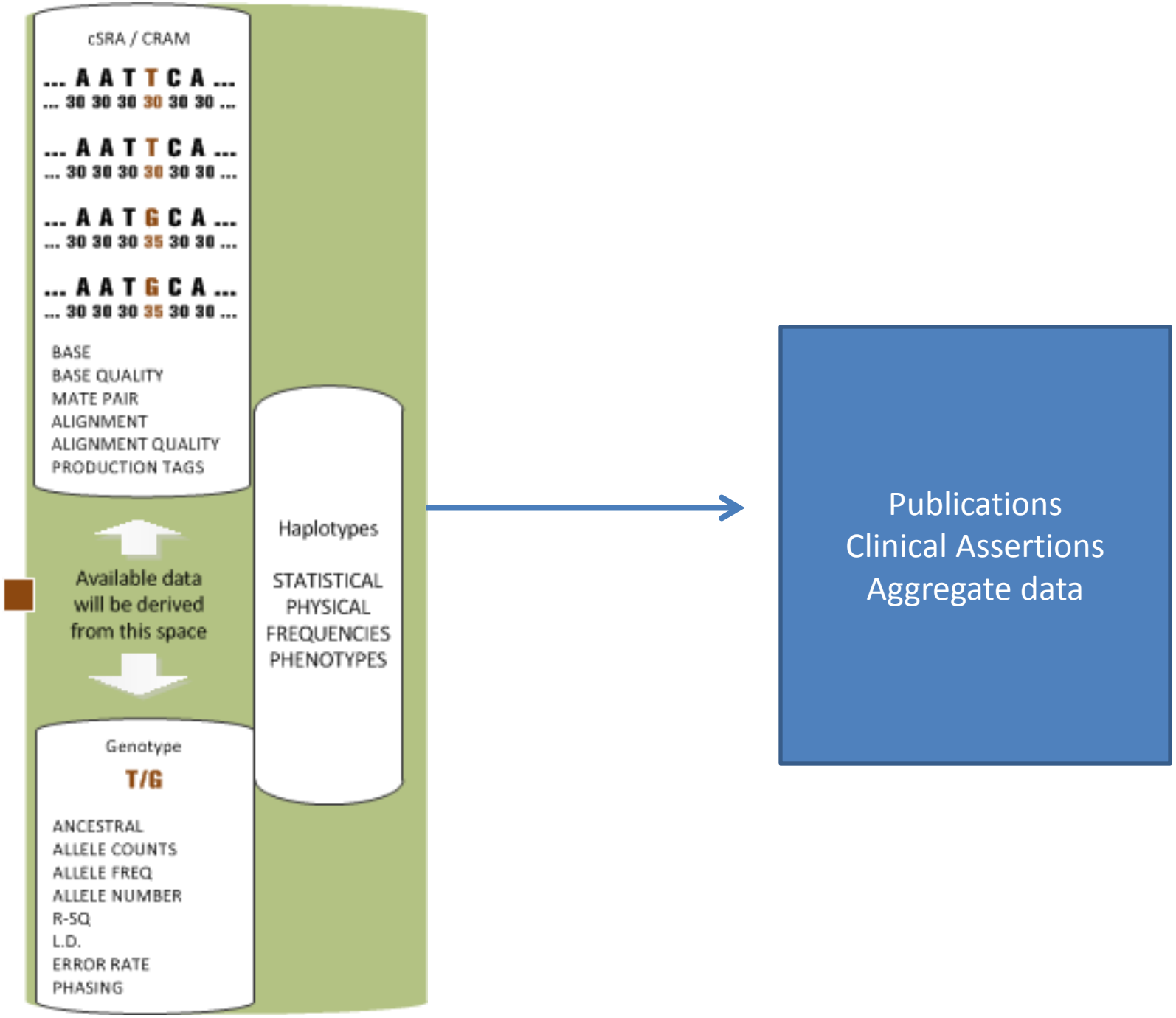
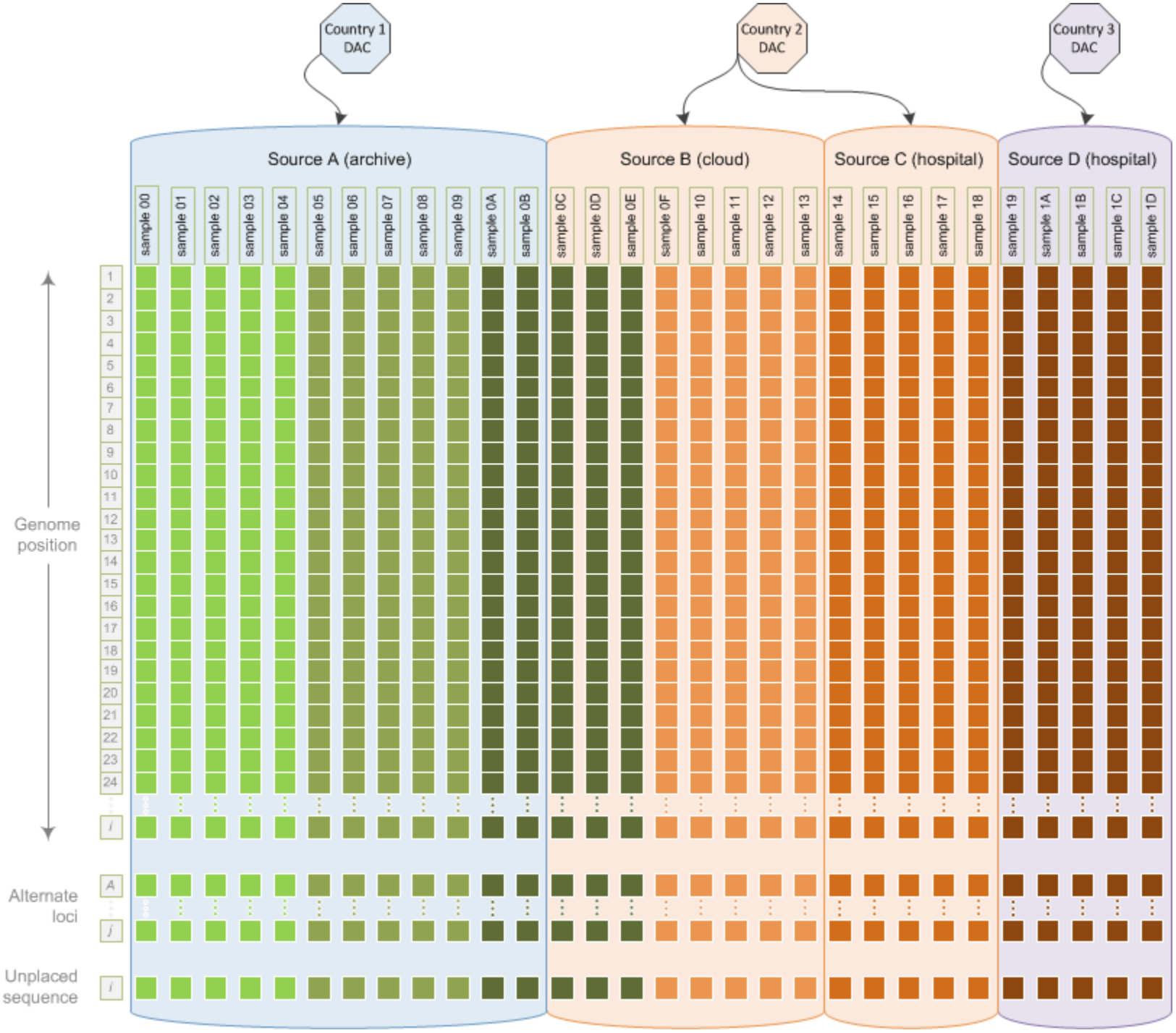
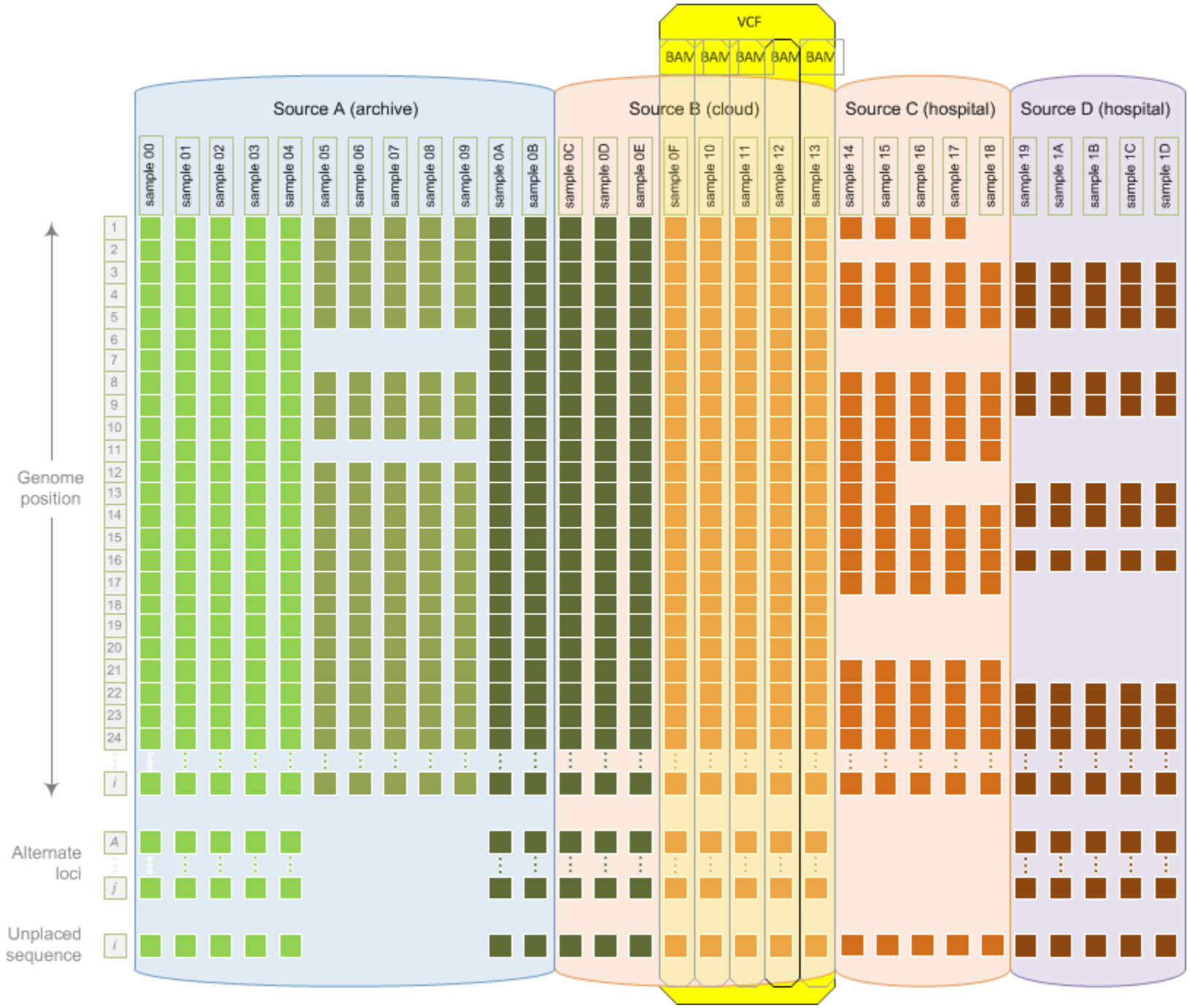
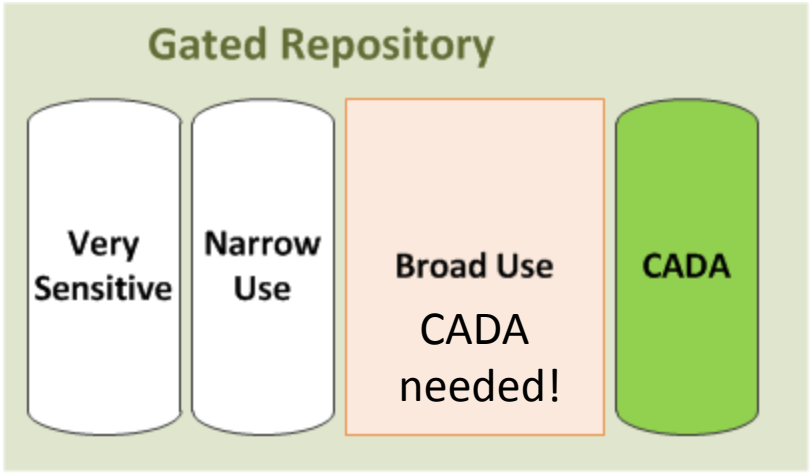
Database, computing capacities and resources


All research participant samples
sample 00
sample 01
sample 02
sample 03
sample 04
sample 05
sample 06
sample 07
sample 08
sample 09
sample 0A
sample 0B
sample 0C
sample 0D
sample 0E
sample 0F
sample 10
sample 11
sample 12
sample 13
sample 14
sample 15
sample 16
sample 17
sample 18
sample 19
sample 1A
sample 1B
sample 1C
sample 1D
Genome position
Alternate loci
Unplaced sequence
1D,10
cSRA / CRAM
... A A T T C A ...
... 30 30 30 30 30 30 ...
... A A T T C A ...
... 30 30 30 30 30 30 ...
... A A T G C A ...
... 30 30 30 35 30 30 ...
... A A T G C A ...
... 30 30 30 35 30 30 ...
BASE
BASE QUALITY
MATE PAIR
ALIGNMENT
ALIGNMENT QUALITY
PRODUCTION TAGS
Available data will be derived from this space
Genotype
T/G
ANCESTRAL
ALLELE COUNTS
ALLELE FREQ
ALLELE NUMBER
R-SQ
L.D.
ERROR RATE
PHASING
Haplotypes
STATISTICAL
PHYSICAL
FREQUENCIES
PHENOTYPES

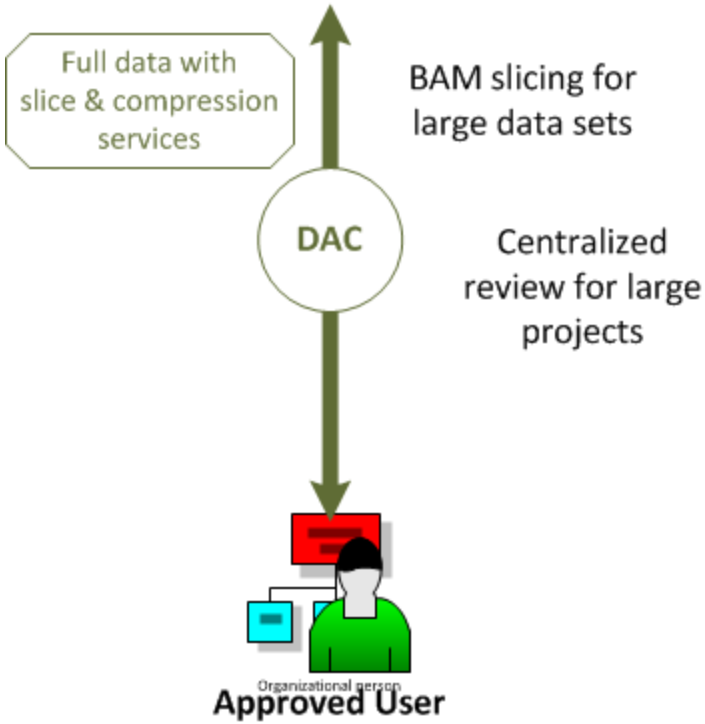

cSRA / CRAM
... A A T T C A ...
... 30 30 30 30 30 30 ...
... A A T T C A ...
... 30 30 30 30 30 30 ...
... A A T G C A ...
... 30 30 30 35 30 30 ...
... A A T G C A ...
... 30 30 30 35 30 30 ...
BASE
BASE QUALITY
MATE PAIR
ALIGNMENT
ALIGNMENT QUALITY
PRODUCTION TAGS
Available data will be derived from this space
Haplotypes
STATISTICAL
PHYSICAL
FREQUENCIES
PHENOTYPES
Genotype
T/G
ANCESTRAL
ALLELE COUNTS
ALLELE FREQ
ALLELE NUMBER
R-SQ
L.D.
ERROR RATE
PHASING
Publications
Clinical Assertions
Aggregate data


Country 1 DAC
Country 2 DAC
Country 3 DAC
Source A (archive)
Source B (cloud)
Source C (hospital)
Source D (hospital)
sample 00
sample 01
sample 02
sample 03
sample 04
sample 05
sample 06
sample 07
sample 08
sample 09
sample 0A
sample 0B
sample 0C
sample 0D
sample 0E
sample 0F
sample 10
sample 11
sample 12
sample 13
sample 14
sample 15
sample 16
sample 17
sample 18
sample 19
sample 1A
sample 1B
sample 1C
sample 1D
Genome position
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
i
Alternate loci
A
j
Unplaced sequence
i

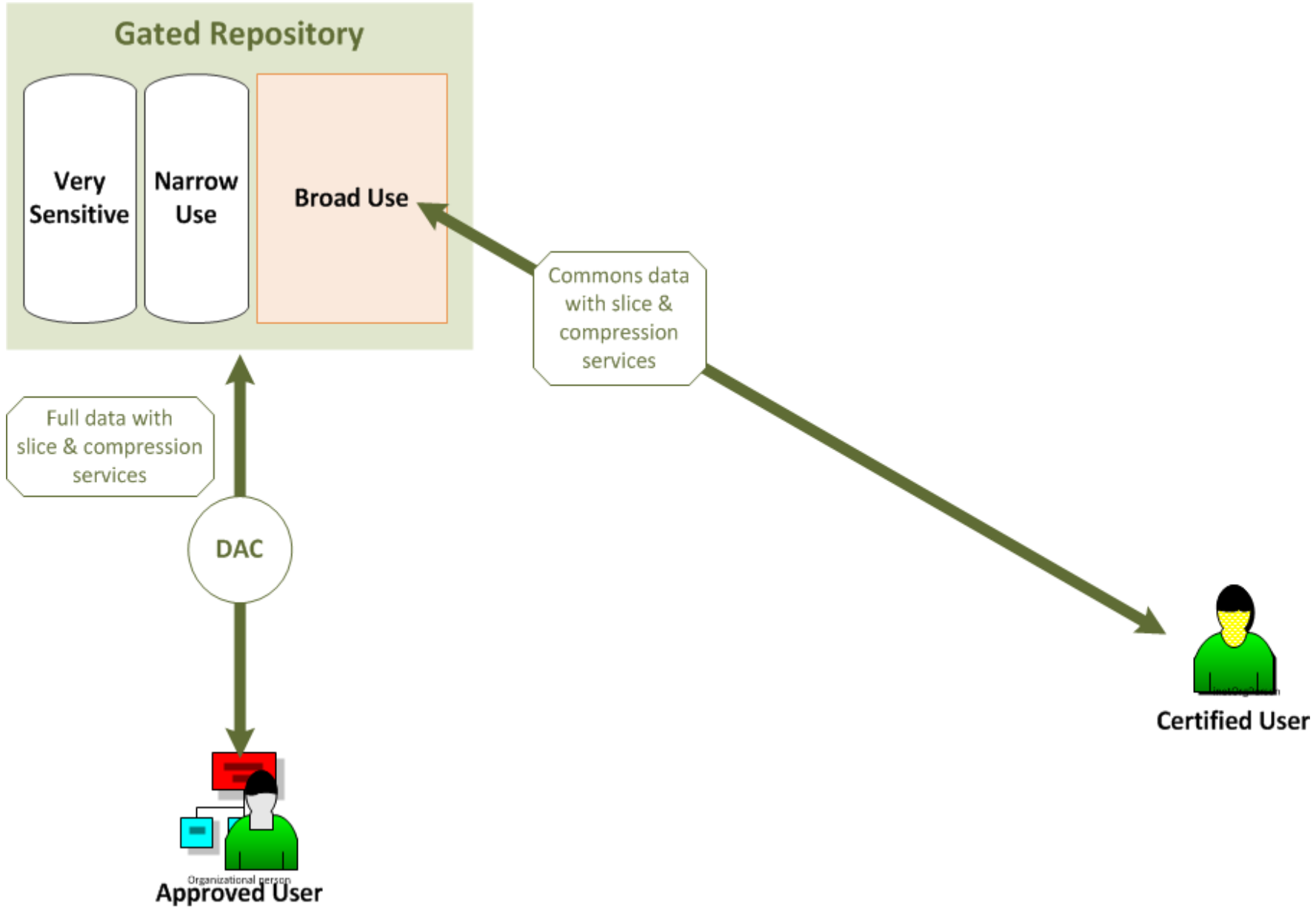

VCF
BAM
BAM
BAM
BAM
BAM
Source A (archive)
Source B (cloud)
Source C (hospital)
Source D (hospital)
sample 00
sample 01
sample 02
sample 03
sample 04
sample 05
sample 06
sample 07
sample 08
sample 09
sample 0A
sample 0B
sample 0C
sample 0D
sample 0E
sample 0F
sample 10
sample 11
sample 12
sample 13
sample 14
sample 15
sample 16
sample 17
sample 18
sample 19
sample 1A
sample 1B
sample 1C
sample 1D
Genome position
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
i
Alternate loci
A
j
Unplaced sequence
i

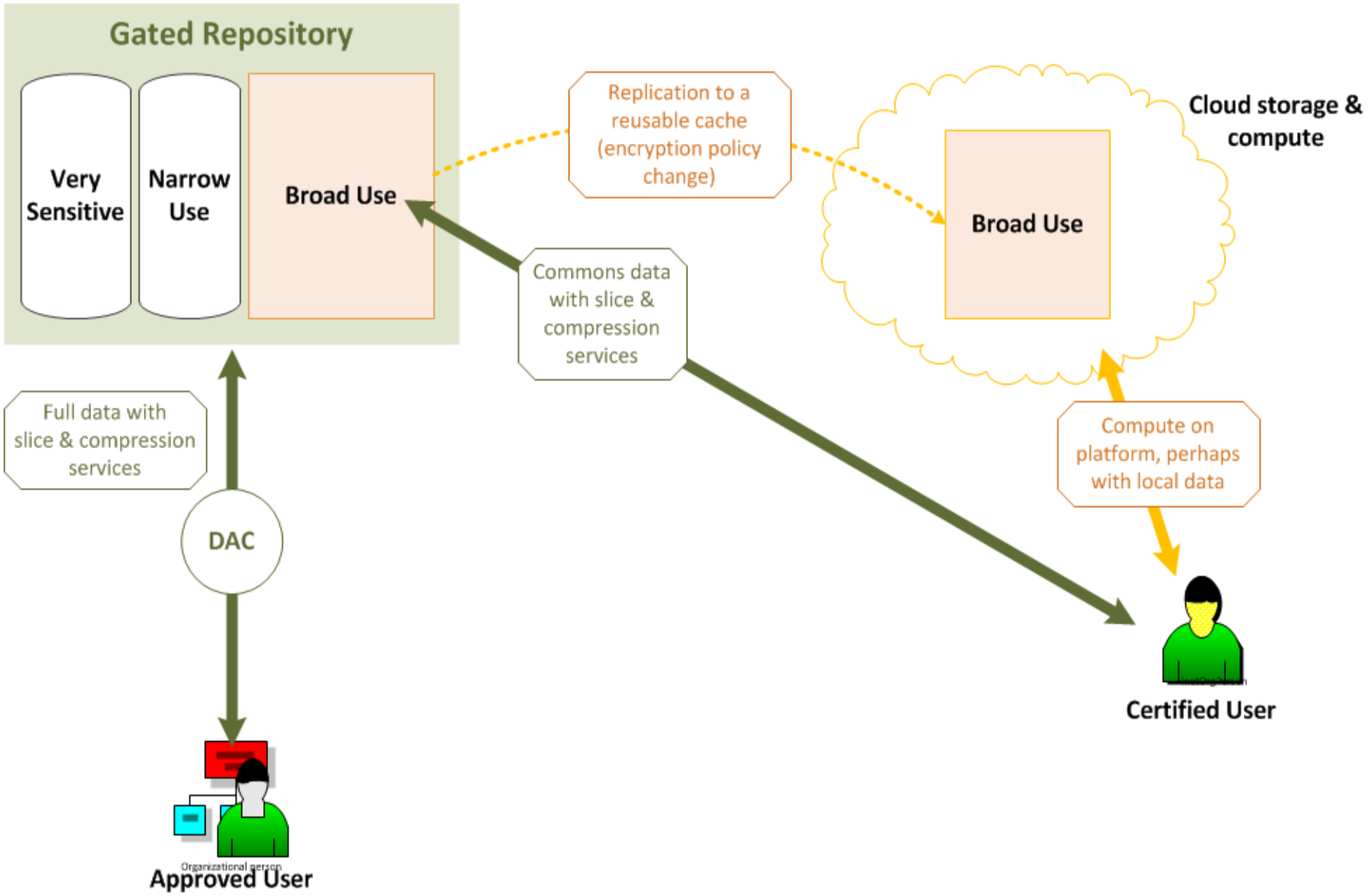


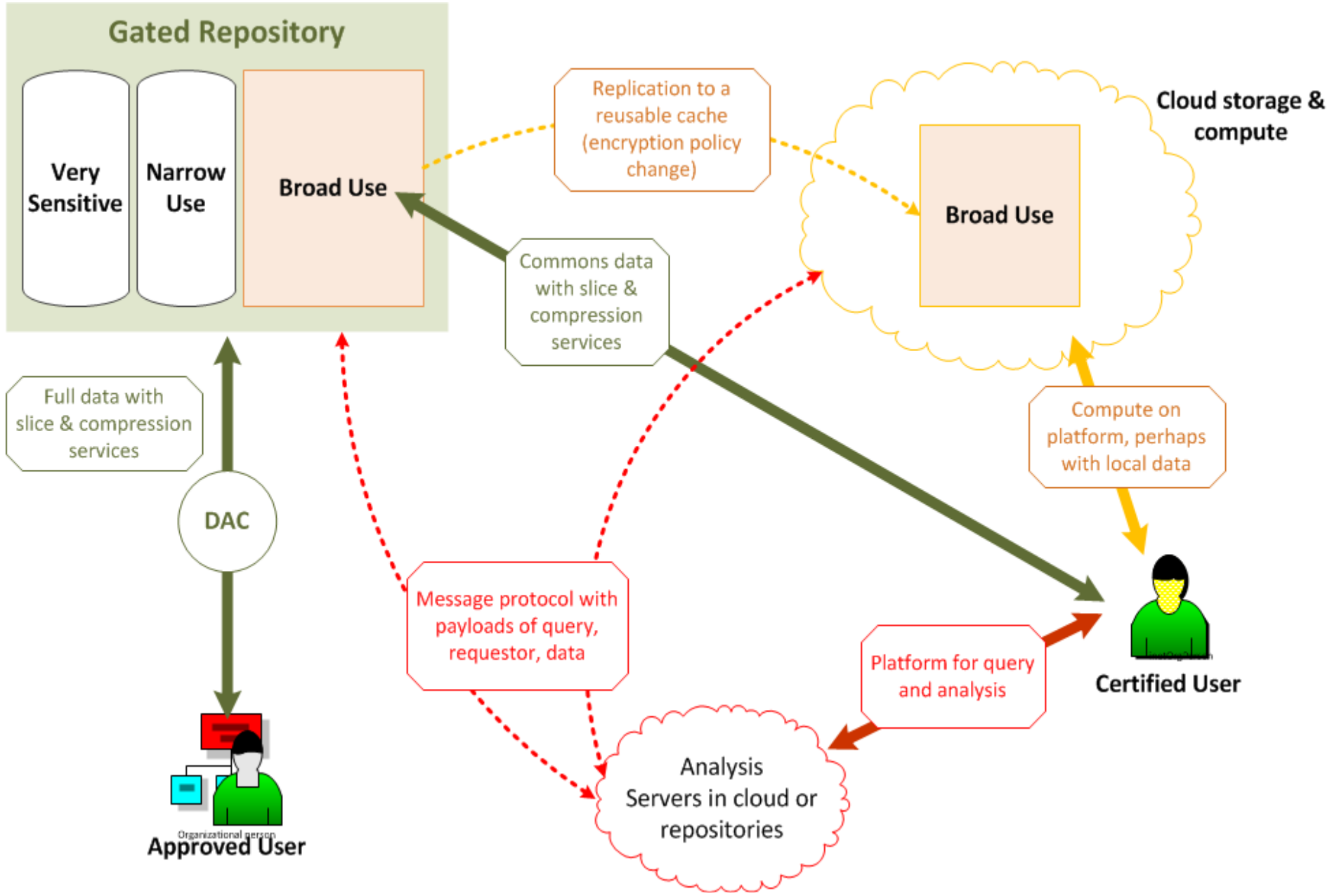
Adding combined aggregate data analysis sets for expedited access

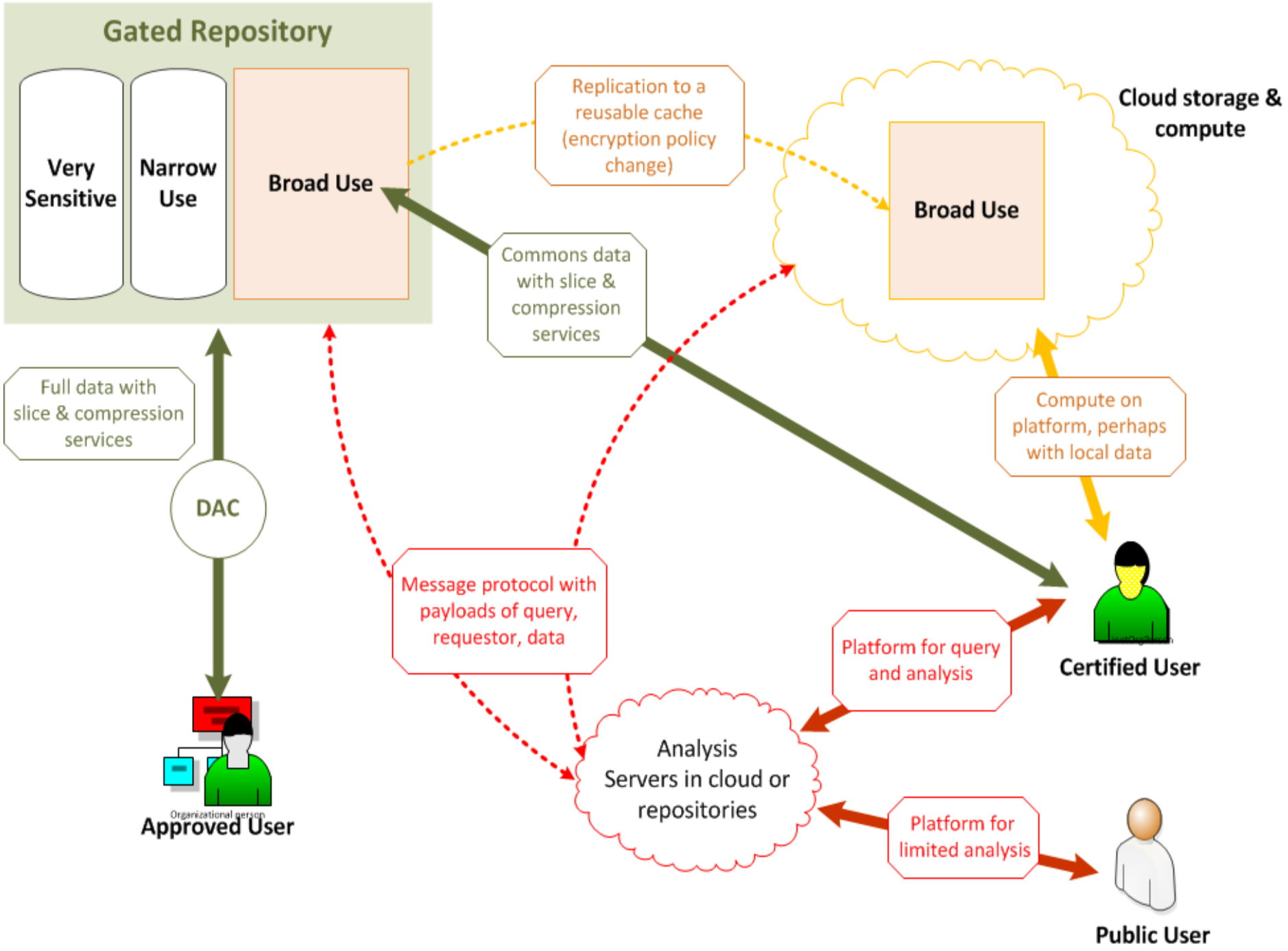


This can be done in July and ready for the fall if recomputes for Broad Use GRU sets are prepared!


Gated Repository
Very Sensitive
Narrow Use
Broad Use
Commons data with slice & compression services
Full data with slice & compression services
DAC
Organizational person
Approved User
Certified User


Gated Repository
Very Sensitive
Narrow Use
Broad Use
Replication to a reusable cache (encryption policy change)
Cloud storage & compute
Broad Use
Commons data with slice & compression services
Full data with slice & compression services
DAC
Compute on platform, perhaps with local data
Certified User
Approved User


Gated Repository
Very Sensitive
Narrow Use
Broad Use
Replication to a reusable cache (encryption policy change)
Cloud storage & compute
Broad Use
Commons data with slice & compression services
Full data with slice & compression services
DAC
Compute on platform, perhaps with local data
Message protocol with payloads of query, requestor, data
Platform for query and analysis
Certified User
Analysis Servers in cloud or repositories
Approved User


Gated Repository
Very Sensitive
Narrow Use
Broad Use
Replication to a reusable cache (encryption policy change)
Cloud storage & compute
Broad Use
Commons data with slice & compression services
Full data with slice & compression services
DAC
Compute on platform, perhaps with local data
Message protocol with payloads of query, requestor, data
Platform for query and analysis
Certified User
Analysis Servers in cloud or repositories
Platform for limited analysis
Approved User
Public User

Elements of a general platform

- Standards for security that permit many individuals to access a source copy.
- Data presentation standards: BAM, VCF, haplotypes, phenotypes, environment
- Message structures for global queries and results. Payload of user properties, question, data permissions
- International coordination for policy & requirements (needs a body)
- Special attention (recalculation) to Research Commons Broad Use studies (or parts of studies)?