

Program of Requirements



Project Napoleon

Owner : Netherlands Forensic Institute
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

TABLE OF CONTENTS

1	DOCUMENT INFO	3
2	INTRODUCTION	6
3	NAPOLEON	8
3.1	PURPOSE OF THE NAPOLEON PRODUCT	8
3.2	FUNCTIONAL REQUIREMENTS	8
3.3	CONSTRAINTS	10
3.4	OPERATIONAL REQUIREMENTS	10
3.5	RELIABILITY REQUIREMENTS	11
3.6	MAINTENANCE REQUIREMENTS	11
3.7	SYSTEM REQUIREMENTS	11
3.8	SECURITY REQUIREMENTS	12
3.9	USABILITY	12
4	BONAPARTE	13
4.1	PURPOSE OF THE BONAPARTE PRODUCT	13
4.2	FUNCTIONAL REQUIREMENTS	13
4.3	CONSTRAINTS	18
4.4	OPERATIONAL REQUIREMENTS	18
4.5	RELIABILITY REQUIREMENTS	19
4.6	MAINTENANCE REQUIREMENTS	19
4.7	SYSTEM REQUIREMENTS	19
4.8	SECURITY REQUIREMENTS	19
4.9	USABILITY	19
4.10	PHASE 2	21
4.11	PHASE 3	22
	APPENDIX ON PEDIGREES	24
	APPENDIX ON ALLELES	25



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

1 Document Info

Document History

Version	Date	Status	Author	Change description
0.1	29-05-2008	Concept	10, 2, e	Draft: after discussions with SNN on 22-05-2008 and 10, 2, e on 26-05-2008. Describes requirements for phase 1, with basic functionality.
0.2	05-06-2008	Concept		After discussion with 10, 2, e
0.3	07-06-2008	Concept		After discussion with SNN on 05-06-2008. Revised by 10, 2, e
0.4	13-06-2008	Concept		Comments SNN processed.
0.5	23-06-2008	Concept		Comments of 10, 2, e (ICMP) processed.
0.6	02-07-2008	Concept		Comments of 10, 2, e processed.
1.0	14-07-2008	Preliminary		Comments SNN processed. This version is the baseline for phase 1; phase 2 will be elaborated in future versions.
1.1	18-09-2008	Concept		Update after discussions with SNN and the creation of the Functional Designs for Bonaparte and Napoleon. Chapter 5 (on Workflow, informative only) has been removed
2.0	07-10-2008	Approved		Approved by SNN and NFI. Reference for phase 1 of the development contract.
2.1	05-01-2009	Preliminary		Update for small changes and errors that came up during development of the basic functionality. Text of BC002 clarified.
2.2	23-02-2009	Preliminary		Changes due to discussions on phase 1 delivery. See minutes of meeting 13-02-2009 for detail.
3.0	20-04-2009	As-built (phase 1)		Approved by SNN and NFI. This version describes the first release of Napoleon and Bonaparte.
3.1	05-06-2009	Concept (phase 2)		Adapted after discussions with SNN on 27-05-2009. For review.
3.2	10-09-09	For approval of phase 2		Adapted after discussions with SNN on 07-09-2009. This version describes the scope-change regarding phase 2 (and phase 3) of the project. This change is for approval by the steering committee.
3.3	10-09-09	For start of stage 3		After discussion in steering committee. mtDNA and Y-STR are considered as indispensable features for Bonaparte phase 2. Scope for phase 2 and 3 therefore adapted tot that.
4.0	17-11-2009	Definition of phase 2		After discussions with SNN and internal NFI. Basis for Bonaparte phase 2. Outline for phase 3.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

Table of References

Reference [ref]	Title	Author	Version	Comments
Bonaparte	Bonaparte: Bayesian networks for victim identification on the basis of DNA profiles	SNN	2.1	
Napoleon	Research Proposal Napoleon (PID)	NFI	3.0	
Dawid	Non-fatherhood of mutation? A probabilistic approach to parental exclusion in paternity cases.	Dawid et al	n.a.	2001 For. Science Int. 124: 55-61

Acronyms en definitions

Acronym	Description
(CODIS) Case ID	CODIS-key: NFI-case ID + KLPD#
Allele	Number of repeating DNA-blocks on a Locus
Batch	Search definition within all DNA profiles in a database, based on a selection of characteristics. Batches are not bound to Projects or Folders
HBS	Humane Biologische Sporen (voormalig BSO: Biologisch Sporen Onderzoek)
CODIS	Combined DNA Index System, application that is licensed by S.A.I.C, USA, under permission of the FBI. Installed version at NFI: 5.7.3.
CODIS DVP	NFI Database containing DNA profiles of missing persons ("Databank Vermiste Personen"), family members of missing persons and of unidentified bodies based on CODIS
CODIS Specimen Category	<ul style="list-style-type: none"> • MP: Missing Person [relation code: A, B] • UP: Unidentified Person [O, P, Q, R] • EK: Elimination known [n.a.] • MP-relative: <ul style="list-style-type: none"> ▪ B(iological)F(ather) [F], BM(other) [M], BC(hild) [C, D, E], BS(ibling) [S, T, U,X] ▪ P(ateral) R(elative) [n.a.], MR [n.a.] ▪ EK: Elimination known [n.a.] • n.a.: Spouse [Z]
CODIS Specimen Data	Data belonging to a Specimen (such as #Loci, profile, alleles etc.)
CODIS Specimen ID	Unique ID for a DNA profile (Database specific, e.g. in CODIS: SIN# + Rol + Aanvraagnummer
DMZ	Demilitarized Zone, physical or logical sub network that contains and exposes an organization's external services to a larger, untrusted network, usually the internet.
DNA profile	See Profile
DT&B	Digitale Technologie & Biometrie
FM	Family Member (Father, Mother, Child, Grandfather, etc.)
Folder	Collection of Projects with a common characteristic, such as "Srebrenica", "Suriname" resp. "DVP" and common match criteria.
GUI	Graphical User Interface
ICIS	Interactive Collaborative Information Systems (http://www.senternovem.nl/bsik/projecten/ict/icis_interactive_collaborative_information_systems.asp)
ICMP	International Commission on Missing Persons (http://www.ic-mp.org)
Individual	Individual person (can be either a UI or a person in a Pedigree, or another Individual Type. An Individual belongs to exactly one Project.
Individual Types	Different types of individuals in Napoleon, such as:



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

Acronym	Description
	<ul style="list-style-type: none"> - UI : Unidentified Individual - MP : Missing Person - PE : Personal Effect - FM : Family Member - EP : Elimination Profile - NI : Non Informative - PR : Profile Removed
Locus	Position of a STR on a chromosome
Mixed profile	DNA profile of a contaminated specimen, containing DNA material of multiple individuals
MP	Missing Person
Mutation model (uniform)	Probability distribution of mutations between allele-pairs of a Locus. Uniform Mutation model (default) assumes that there is a propagation probability P for each allele. The remaining probability (1-P) is distributed evenly over mutation towards one of the other alleles. See Dawid et. Al.
NFI	Netherlands Forensic Institute
NFI#	NFI case number
Null-allele	Also called "silent" allele: a DNA replication artifact due to allele dropout causing the allele to be invisible as a peak in the DNA profile.
PE	Personal Effect (a.k.a. Personal Item), object believed to belong to a Missing Person, such as a toothbrush containing DNA material.
Pedigree	Construction of the mutual relations between missing persons and their relatives.
Profile (AM, PM)	Anonymous DNA profile related to a case. A UI profile is Post Mortem (PM), collected from unidentified remains. An Ante Mortem (AM) profile is either: <ul style="list-style-type: none"> • collected from a family member of a missing person • collected from a personal item, such as a toothbrush of a missing person
Project	Work set within Napoleon, consisting of one or more DNA profiles e.g. linked to the same missing person, Pedigrees or unidentified person. A project consists of : <ul style="list-style-type: none"> • 0 or more unidentified individuals • 0 or more (putative) AM profiles. (A putative AM profile can be a PM profile) • 0 or more (putative) Pedigrees If a Pedigree consists of more missing persons, it could be stored in one Project.
Role	Role in a Pedigree, relative to the MP: (Father, Mother, Child, Sibling, PE, unidentified)
Search Count	Number of times a profile was searched for
Search Set	Set of UIs to be matched. A Search Set is part of a Project.
SIN	Specimen Identification Number, represents a "Stuk van Overtuiging"
SNN	Stichting Neurale Netwerken
SNP	Single Nucleotide Polymorphism
Specimen Type	Also known as Specimen Category. Attribute designating the type of a DNA profile, such as "MP", "BM", "BF", etc.
STR	Short Tandem Repeat, repetitions of tiny pieces of DNA, used for matching because they can vary strongly between individuals
UI	Unidentified Individual or Personal Effect ¹
Virtual profile	Constructed profile, combining two or more partial DNA profiles
XML	Extended Markup Language

¹ A PE is either positively belonging to an Individual (e.g. a tooth from youth) or maybe belonging to an Individual (e.g. a toothbrush, which might also belong to a sibling). In the first case the PE acts as a FM, in the latter it acts as a UI.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

2 Introduction

Napoleon is a system for identification of unidentified victims through screening and matching of their DNA profiles against Pedigrees of Missing Persons' DNA Profiles in large databases based on a Bayesian statistical network analysis approach.

The development of Napoleon will be performed and supported by the following parties:

Netherlands Forensic Institute (NFI)

- Department Humane Biologische Sporen (HBS)
- Department Digital Technology and Biometrics (DTB)

Stichting Neurale Netwerken (SNN)

the Dutch Foundation for Neural Networks, a national network of research groups on neural networks, machine intelligence and affiliated techniques. SNN Nijmegen, the research group [10.2.g](#) will be involved in this project.

International Commission on Missing Persons (ICMP)

- Research lab in Bosnia which is responsible for identifications of victims in former Yugoslavian countries, after Tsunami disaster, Katrina hurricane and after other disasters.

Within the Napoleon research project the aim is to develop new software in which the whole Pedigrees of the relatives of missing persons are used in both the screening and the matching phase. The main advantage of this approach is that the number of false hits will be much lower than conventional methods. The software will consist of database functions, a graphical user interface and screening and matching functionality. The Matching of Missing Persons against its Pedigree's in large databases will be based on a Bayesian statistical network analysis approach. In this way the Napoleon system will improve the quality and continuity of the services delivered by the NFI, even in case of a major disaster. The combination of both the Pedigree Approach and the Bayesian Approach is new in DNA identification work.

Within the Napoleon project two complementary products will be developed:

- Napoleon-product: the Logistics modules to be developed by NFI
- Bonaparte product: the Calculation core to be developed by SNN

These products will be specified in the following chapters. As a rough indication, the domains of both products are sketched globally in the picture below.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

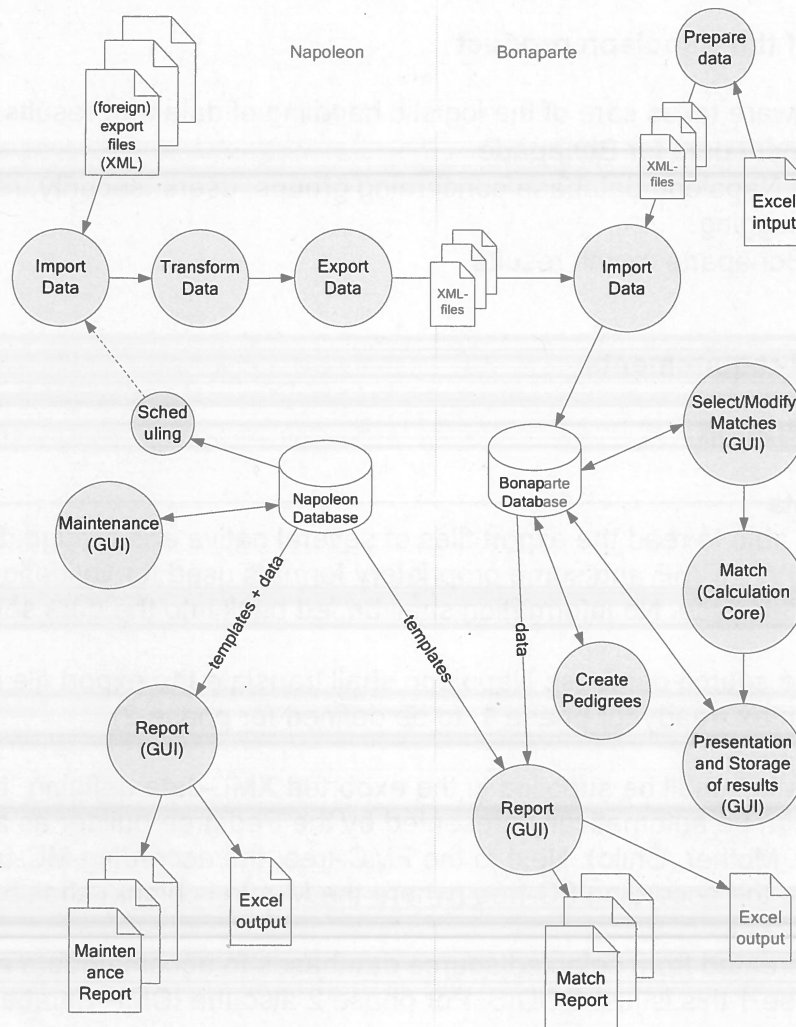


Figure 1: the Napoleon and Bonaparte domains

The project is divided into two phases :

- Phase 1: Autosomal DNA profiles
- Phase 2: Extended profiles

The distinction between these two phases is indicated in the text, where relevant. Phase 1 focuses on good-quality profiles and the general feasibility and usability of the method, whereas phase 2 is concerned with refining and extending the method.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

3 Napoleon

3.1 Purpose of the Napoleon product

The Napoleon software takes care of the logistic handling of data and results:

- Preparing the input data for Bonaparte
- Maintaining the Napoleon database concerning groups, users, security, report-templates and logging.
- Reporting the Bonaparte match results

3.2 Functional requirements

3.2.1 Data Preparation

[NF001] Import data

Napoleon shall be able to read the export files of several native and foreign databases. (phase1: CODIS DVP, ICMP and some proprietary formats used for validation purposes). This is a one-way interface. No information shall be fed back into the data sources.

Independent on the source database Napoleon shall translate the export file into a XML-file with a standard syntax (ready for phase 1, to be defined for phase 2).

Basic Role information shall be supplied in the exported XML-data defining "basic" Pedigrees which can be automatically presented by the Pedigree builder as a first guess (FMC-tree: Father, Mother, Child). Next to the FMC-tree, the according MC-tree (where the F-role is blank) and the according FC-tree (where the M-role is blank) shall be supplied.

Daily exports are created from selected source databases to accommodate automatic matching. For phase 1 this is just CODIS. For phase 2 also the ICMP-database should be equipped with this.

Napoleon associates all imported profiles with at least:

- Individual Type
- Project
- Folder

This associated information will be used by Bonaparte to configure matches.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

Possible Individual Types:

Individual Type	Reference or Target	DNA present	ID known	Description
FM	Reference	✓	✓	Family Member (CODIS: 1 st degree: F, M, C, S, in Napoleon: no restrictions on degree)
UI	Target	✓	x	Unidentified Individual
PE	Target	✓	possibly	Personal Effect (either UI (no ID) or FM (ID))
MP	Target	x	✓	Missing Person
EP	Reference	✓	✓	Elimination Profile
NI	Reference	x	✓	Non Informative (family member needed to complete the pedigree)
PR	Reference	x	✓	Profile Removed (e.g. for legal reasons)

3.2.2 Maintenance

[NF002] Format flexibility

Identification fields used in database structures, such as Specimen ID numbers, must be flexible so that data from various sources can be processed.

Minimal sources are CODIS DVP, the ICMP database and Test databases.

[NF004] DNA-profile manipulation

Obsolete, contained in [BF003]

[NF005] Static Data Configuration

Obsolete, contained in [BF002]

[NF006] Archiving (phase 2)

Obsolete, joined with [BF009]

3.2.3 Reporting

[NF007] Reporting features

Napoleon can deliver the following user specified Reports using data from the Napoleon internal database:

1. Match Reports (NFI)
2. Match Reports (ICMP)
3. Match Reports (ad-hoc) (phase 2)
4. Deleted-Reports, when a case is (virtually) deleted from Napoleon

[NF008] Report archeology

Obsolete, contained in BO003]



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

3.3 Constraints

[NC001]

The technical environment in which Napoleon is developed should be MySQL/Java and Flexible or Open source platforms, as far as performance requirements permit this.

[NC004]

The XML-conversion of databases to be imported should be compatible to the DNA profile structures used by CODIS DVP and the ICMP (phase 1) as well as Excel Validation data in ad-hoc formats. In a later stage other conventions may be added.

[NC006]

Napoleon must be validated with a proper test-set to be defined by NFI and ICMP.

Note: DNView cannot be used as a formal (commercial) validation reference (prohibited by Brenner). Probably Familias will be used for this purpose.

[NC007]

The interfaces with databases containing DNA profiles, or exports from databases, are read-only and limited to DNA profiles with their reference information, such as case-id etc. Napoleon does not feed back any information into its source databases.

[NC008]

There is no direct interface between Napoleon and its source databases.

[NC009]

Linkage (relations between alleles) is not applied within Napoleon, except possibly for phase 2 when SNP is included.

3.4 Operational requirements

[NO001] Client-server and standalone mode

Napoleon will be able to run standalone (phase 2) on a commodity laptop as well as in a network environment with multiple users (phase 1) in a client-server architecture.

The maximum number of concurrent users is 20 Forensic Researchers and/or 50 Data-enterers (might be needed in the event of a calamity)

[NO002] Logging of actions

All searches and decisions by the user shall be logged, specifying time, user and action.

[NO003] User administration

A user-administration must be set up. This administration contains the known users and their according rights to operate Napoleon and provides Role Based Access Control.

The following User Roles and according privileges are identified:



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

User Role	Privileges	Comments
Administrator	all	System administrator. Includes the SNN maintenance user)
Data Enterer	Basic maintenance	Data Entry (profiles, Pedigrees, etc.)
Forensic Researcher	Matching, Basic maintenance	NFI personnel or ICMP-personnel

3.5 Reliability requirements

[NR001] Backups

There shall be periodic automatic backups so that minimal work will be lost in case of problems. After recovery at most one hour of work can be lost. The backup only concerns the internal Napoleon database. An automatic permanent backup is made daily.

3.6 Maintenance requirements

[NM001] Compatibility

Napoleon's interface conversion modules must keep up with:

- new releases of source databases
- new statistical models
- new DNA kits (LociSets)

[NM002] Export

For each selected database source, Napoleon shall supply export functionality, as far as not contained in the source database itself.

[NM003] Use of 3rd party products

All 3rd party products applied in Napoleon should be acquired from reliable suppliers, offering transparent conditions for usage, upgrade and maintenance.

[NM004] Documentation

Clear design documentation and manuals are needed for the use and maintenance of Napoleon.

3.7 System requirements

[NS001] Interfacing with databases

Napoleon will have database interface modules allowing it to read exports from CODIS DVP, ICMP and other foreign DNA-databases.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

3.8 Security requirements

[NX001] Authorization

Only authorized users may gain access to Napoleon through a login procedure. Napoleon starts up with an authorization-dialog. After a configurable period of inactivity the authorization dialog reappears, so that the user has to log in again.

[NX002] Privacy

From NFI's viewpoint it is preferred that no privacy sensitive information is contained within Napoleon.

Bonaparte should nevertheless *be capable* of containing privacy sensitive information. This information might be very useful in checking the consistency of matches, and other potential users may require the use of such data.

Specifically the following text fields must be accommodated:

- Name
- Date of Birth
- Gender
- Ethnical group (population)
- Residence
- Finding place of sample
- "Missing since" date

plus some future-use fields.

The use of each field must be configurable. If a field is not configured for use, it is invisible to the user and contains no data (default setting for NFI-use).

If privacy sensitive information is used and stored, it must be encrypted such that it can only be reconstructed by using Napoleon (phase 2). Napoleon shall be run isolated from networks that contain privacy sensitive information, e.g. in a DMZ.

[NX003] Secure connections

No insecure open internet connections may be applied.

3.9 Usability

[NU001] Look and Feel

The GUI is designed to be intuitive, conforming to universal GUI-guidelines.

[NU002] User Profiles

Napoleon shall support user-profiles containing preferred settings, etc.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

4 Bonaparte

4.1 Purpose of the Bonaparte product

The project “Bonaparte” will deliver a system for kinship analysis in DNA-files as described in the ICIS valorization proposal [Bonaparte].

This system, the computational core of Napoleon, will be referred to as “Bonaparte”.

Bonaparte consists of 3 elements:

- Computational core
- GUI
- Interface (database, XML)

4.2 Functional requirements

4.2.1 Computational Core

The computational core performs the screening and matching. Matching can be done either automatically² or manually.

Automatic Matches consist of (“UI vs. anything”):

- direct match UI vs. UI
- direct match UI vs. PE
- kinship match UI vs. MP (Pedigree)

Manual matches can consist of (“anything vs. anything”):

- direct match UI vs. UI
- direct match UI vs. PE
- kinship match UI vs. MP (Pedigree)
- kinship match PE vs. MP (Pedigree)

Note that matching a MP is in fact matching against the MP’s Pedigree (there is no DNA-profile of the MP self).

Note also that there is no direct matching against references (only kinship matching)

The computational core shall support the following features:

[BF001] Elementary matching

Elementary matching options (Search Set vs. Batch), where x, y are numbers [1,N]

Direct matches:

1. x UI vs. y UI (to detect doubles)

² Automatic matches are not explicitly triggered by the user but must be activated after addition of profiles by the external system(s), or the countdown of a timer. They are performed in the background.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

2. x UI vs. y PE

Kinship matches:

3. x UI vs. y Pedigrees; match UI-profiles into (putative) Pedigrees
4. x UI vs. y FM (1:1) to detect parents, children and siblings in Pedigrees, with the same role as in the current Pedigree: this means that for the screening to detect parents and children only the family members with the roles F, M and C are involved. To detect siblings only the family members with the role S are involved.
5. (phase 2) x UI vs. y UI (to detect parents, children and siblings)

[BF002] Network builder

Build Bayesian network family – per locus

1. Paternal/Maternal inheritance
 - a. Mutation model (such as from Ayres or Brenner.
In phase 1 only the uniform mutation model will be applied)
2. Genotype observation
 - a. Alleles
 - b. Multiple genotypes with their according probabilities³
3. Population allele frequencies
4. Theta-correction (phase 2, exactly or approximately implemented)
5. Rare alleles and Size Bias Correction (phase 2, minimal count, via λ_{common} λ_{rare} λ_{new} , see Appendix on alleles),
6. Null-alleles

It is acknowledged that applying complex models or corrections will impair the system performance. Performance benchmarking on screening will be done with simple models and corrections, whereas more complex models and corrections will only be applied for validation purposes (in phase 2).

(SNN will investigate the performance effects of the selection of corrections and models)

4.2.2 GUI

The GUI shall allow the user to

- enter configuration data and selections for the computational core
- perform basic manipulations on Profiles, Folders and Projects
- manipulate Pedigrees

[BF003] Profile Management⁴

The following operations are possible with profiles. Reference is the Napoleon internal database. Adapted profiles are never fed back into the source-database.

For profiles constructed through the Napoleon GUI:

- 1) Add a profile (NB: an artificial profile must be clearly marked as being artificial)

³ Possibly to be related to peak-heights in a later stage (phase 2 or later). NB mixed profiles are excluded from the computational core in both phase 1 and 2. However the data-storage should be prepared for storing mixed profiles which may be dealt with in future versions.

⁴ The system will compute with at most one (maybe virtual) profile per individual.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

- a) Family Member (FM)
 - b) Personal Effect (PE)
 - c) Victim (UI)
 - d) Body Part (UI)
- 2) Edit/Change a profile (NB: an adapted profile must be clearly marked as being artificial)
- a) Population⁵
 - b) Pedigree reference(s)
 - c) Alleles
 - d) Add Loci

For external profiles:

- 3) Import profiles from an external source (XML or Excel)

For any profile:

- 4) Export profiles to an external file (XML-interface)
- 5) (phase 2) Construct a virtual profile (e.g. add up of profile parts supposedly belonging to the same victim, or making a test-profile)

[BF004] Construction of Pedigrees

Bonaparte can import pedigrees in a specified XML format

In addition, the Bonaparte GUI offers a tool to construct Pedigrees graphically.

Pedigrees consist of squares (males) and circles (females), where the color represents the status being either

- “no DNA sample” = checked square/circle,
- “reference sample” = blank square/circle,
- “missing person” = red square/circle,

Pedigrees may be composed manually by dragging Individuals that belong to the Project into the Pedigree.

Basic role information is supplied by the imported XML-data and is displayed automatically.

It shall be possible to define relations within the Pedigree such as: there is a 10% chance that a particular child is not of the supposed biological father.

Note that a Pedigree can “evolve” by dragging Individuals onto its positions. A PM profile may act as an AM reference profile when it is dragged onto a reference-position (it will always remain a target however). See also “Appendix on Pedigrees”.

One project may contain several (putative) Pedigrees. A (putative) Pedigree can be excluded from use in Matching.

⁵ For phase 1 Population is coupled to a Match, not to an individual profile. In later stages the default Population-value of a Match might be overruled by a profile-bound Population-value. For phase 1 however, this field is unused.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

[BF005] Pedigree Consistency Check

Upon user request the system computes a consistency measure of a Pedigree and presents the number of allele mismatches in parent(s)-child relations contained in it. Judgment of the overall consistency is left to the user, who may decide to accept, reject and/or store the Pedigree, or adapt the Pedigree until the Pedigree is considered sufficiently consistent.

[BF006] Match Preparation

The Bonaparte GUI will list the automatically imported data/new projects and earlier defined projects. Data affected by automatic searches will be distinguished from unaffected data.

A (set of) NFI-case(s) may be given a name so that the project can be easily found again. New imported data: user should define a project name and may add Pedigree(s).

For the match preparation the user should define the Search Set

Now the list with Individuals belonging to the selected project is presented in a list, together with the attribute "Individual Type", for the Search Set.

Next a Batch can be defined.

The following criteria can be used to create a batch:

- Folder Name
- Project name
- Individual Type
- Individual

The created batch is the matching set for the Search Set; it cannot be saved as a separate entity within a Folder. There can be selectable predefined Batches.

The user may now select the option "match", to start the Match-Setup dialog.

[BF007] Match Parameters

Match Parameters can be applied on two levels: Folder and Project.

For automatic searches the Folder-level Match Parameters are applied

For manual searches the Project-level Match Parameters supersede the (default) Folder-level Match Parameters. The following Match Parameters must be set:

Direct Matches:

- 1) Population Type
- 2) Subpopulation correction (⊖-correction, phase 2)
- 3) Size bias correction/rare allele model (phase 2, as modeled with λ_{common} λ_{rare} λ_{new} , see Appendix on alleles)
- 4) Prior odds (can be set afterwards at Match Details)

Kinship Matches (extra parameters compared to direct matches):

- 5) Mutation model (or specify a customized mutation model)
- 6) Model for rare alleles (phase 2)
- 7) Error model (phase 3)



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

On Folder-level, where the default Match Parameters are specified, it also has to be specified against which other Folders and Elimination Projects (see BF011]) the Folder has to be automatically matched.

The same remark with respect to performance as in [BF001] applies.

[BF008] Match and Presentation of results

The results of the match are presented to the user.

Minimal content of the Hitlist (layout is to be determined):

- 1) Date, Issuer, Project, Match strategy summary, Case ID
- 2) List of Hits 1 { UI/MP | Reference | LR | #Loci } N

The Hitlist may be filtered as follows:

- Thresholds, such as LR
- Statistics type
 - LR (likelihood Ratio)

Detailed information will be shown on selection of a Hit:

- Date, Issuer, Project, Search Set, Batch, Match strategy summary, Case ID, Folder
- Pedigree
- LR per Locus
- DNA profiles
- Posterior Odds (upon entrance of prior odds)

The detail report can be exported and printed.

The DNA profiles of the Set can be exported for comparison purposes.

Matches can be confirmed by the user via the "Match status". The Match status is initially "Queued", and can be changed into "Pending" (preliminary confirmed) or "Rejected".

The merging of Individuals as a result of a Match status is done outside the scope of Napoleon.

[BF009] Creating Folders and Projects

Projects are formed around one or more MPs (if those MPs are present in the same Pedigree) and/or one or more UIs. Projects must be grouped in Folders.

Both Folders and Projects can be created new.

After some time a Folder or Project(s) may become obsolete so that the user does not want to see these Folders/Projects in selection lists anymore.

Such Folders/Projects can be archived. This means that they are removed from the current workspace, but can still be recovered and put back in the current workspace.

This feature can be used in case of audits or re-opening old cases.

Projects cannot be physically deleted.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

[BF010] Refresh Folders and Projects

Periodically (configurable) the contents of the export of source database may have changed, e.g. profiles were added. These changes may affect Projects and/or Folders. Napoleon should warn the user for this and present the list with possibly affected Projects/Folders.

All new data should be either:

- Added to an existing Project
- Put into a new Project

[BF011] Elimination of Profiles

Profiles can belong to persons that may have contaminated samples in the process, such as policemen, lab-personnel, etc.

Such Profiles are clustered in Elimination Projects, which are on their turn contained in the so-called Elimination Folder.

The automatic search feature tags Hits with Profiles from Elimination Projects that are coupled to Folders (see BF007)).

4.2.3 Interface

Exchange of data between Napoleon and Bonaparte is implemented through the common internal Napoleon database (see Figure 1) and by importing XML-format files.

4.3 Constraints

Napoleon's constraints (paragraph 3.3) NC001, NC006 and NC009 apply to Bonaparte as well.

[BC001]

Mixed profiles are excluded from the computational core.

However the data-storage should be prepared for storing mixed profiles.

[BC002]

The Bonaparte database stores percentages (phase 1:one genotype with 100% percentage). Peak-heights will not be stored in phase 1.

Example: P(a,b) Q(b,b), R(c,d) is used as:

- Phase 1: 100% probability on (a,b)
- Phase 2:
 - $P/(P+Q+R)$ probability on (a,b)
 - $Q/(P+Q+R)$ probability on (b,b)
 - $R/(P+Q+R)$ probability on (c,d)

4.4 Operational requirements

Napoleon's operational requirement NO001 applies for Bonaparte as well.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

[BO001] Performance

The time needed to perform 100 UI vs. 100 basic Pedigrees (F + M + C = MP) matches shall not exceed 1 minute, applying a uniform mutation model.

[BO002] Operational Quality

The operational quality of Bonaparte shall be verifiably better than with the currently used method, in particular in the case of many UI's and MP's with kinship information. Verification, validation and quantification (false hits, true hits, operational processing times) still has to be worked out (a.o. dependent on external factors such as data quality).

[BO003] Logging of data manipulations

Bonaparte logs all data manipulations and stores them in a database such that the database can be rolled back in time. Note that user actions are not logged in Bonaparte and are not to be stored in the database⁶.

4.5 Reliability requirements

[BR001] Search count

The number of searches performed on a specimen can be counted in the database.

4.6 Maintenance requirements

Napoleon's maintenance requirements (paragraph 3.6) NM001, NM003 and NM004 apply to Bonaparte as well.

4.7 System Requirements

4.8 Security requirements

Napoleon's security requirement (paragraph 3.8) NX003 applies to Bonaparte as well. In particular, Bonaparte will not be provided privacy sensitive information.

4.9 Usability

Napoleon's usability requirements (paragraph 3.9) apply to Bonaparte as well.

[BU001]

Bonaparte shall support the use of user-profiles, such that Napoleon user-preferences connected to a user-profile (see requirement NU002) are automatically applied.

[BU002]

⁶ This would lead to unmanageable amounts of data. User actions might however be reconstructed, if needed, from the access-logs of the web-server.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

The parameters used in the model shall be transparent to the user and adaptable where relevant.

[BU003]

A user training shall be part of the delivery.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

4.10 Phase 2

This paragraph describes extra requirements for phase 2 of the Napoleon / Bonaparte product.

1. Expand the Bayesian network with:
 - 1.1. Y-chromosomal data: storage of Y-STR data, direct comparisons with Y-STR data. It will be possible to assign a Y-chromosome haplotype to a male individual. Bonaparte will be able to compute the minimum number of mutations that are needed in the pedigree when a MP is substituted with a UI, minus the minimum number of mutations that are needed when the MP is not replaced with a UI. This number will be used as match result. When a manual match is performed, the user may choose whether or not Y-STR matching is done. The full list of match results can be filtered according to the Y-STR match results.
 - 1.2. mtDNA data: accommodation of storage as an additional profile. The match process is based on the Levenshtein distance (minimum number of deletions, insertions or substitutions of a single character to transform one string into another) between the UI's profile and the profiles in the pedigree.
 - 1.3. Mutation model. The parameters for the uniform mutation model will be made gender- and locus-dependent. In addition, investigation of the benefits of a more refined mutation model.
2. BF001: x UI vs. y UI (indirect matching to detect parents, children and siblings)
Automatic and Manual DVI (with invisible pedigrees)
3. Filter option to segregate Match-results in e.g. 'Screening-matches' (automatic, manual) and 'Matching-matches' (manual) to help avoid flooding the user with match results. A functional design of how this should look like will be prepared by NFI.
4. Permanent deletion of Projects/Folders/Profiles⁷
5. Requirement expansions from phase 1 ([referenced requirement]):
 - 5.1. [NF001] Accommodation of ICMP-database
 - 5.2. [NF007] Ad-hoc match reports
 - 5.3. Extensions to export / import Y-STR and mtDNA data



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

4.11 Phase 3

This paragraph describes extra requirements for phase 3 of the Napoleon / Bonaparte product.

This is a yet-to-be-decided phase. This paragraph enumerates the wish list of features to be considered for such a phase. Partly it contains postponed features that were considered not very useful right now, but probably later on. The other part contains features that were identified as useful during the development of phase 1, but were omitted in this Program of Requirements earlier on.

1. Expand the Bayesian network with:
 - 1.1. Implementation of Theta correction models (BF002) It will be possible to incorporate the theta correction into the calculation of the Likelihood Ratio; the user will be able to choose theta.
 - 1.2. Kinship analysis module apart from the DVI module: In this module, various hypotheses, each containing one or more pedigrees, can be tested. The likelihoods of all hypotheses can be calculated and scaled w.r.t. any one of the hypotheses in order to compute likelihood ratios. In this module theta correction can be applied.
 - 1.3. Y-chromosomal data.
Computations with Y-STR data in line with recent developments
 - 1.4. mtDNA data:
Computations with Y-mtDNA data in line with recent developments
 - 1.5. SNP-markers.
Computations with SNP-data. Linkage between SNP-markers.
 - 1.6. Mutation model [Dawid] (BF002). *The user will be able to choose between several mutation models. This choice can be done per locus and per gender.*
[It must be possible to select the most adequate mutation model from a list of models.]
 - 1.7. Familial Searching module⁷. In this module, the user can select a sample S and a folder F containing reference samples. The user can choose to compute LR's for parent-child and/or sibling relationships (vs unrelated) between S and the profiles in F. In addition, the user can specify thresholds above which the LR must lie in order to be reported.
 - 1.8. STR peak heights
 - 1.9. Detailing of Minimal count method and Rare allele method (BF002, BF007)
The user may choose how to deduce an estimate of allele frequencies from the population database: by using a minimum count method or a Bayesian method. According to the minimum count method, if the database contains N alleles then the allele frequency used in the calculation of the Likelihood Ratio is the maximum of k/N and the database frequency, where the user may choose k between 3 and 5. According to the Bayesian method, the frequency of allele i is estimated as $(N_i + L_i)/(N + L)$ where N_i is the number of type i alleles in the databases, L_i is

⁷ Possible privacy restrictions still to be examined by NFI.



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

chosen by the user and depends on the allele type (common, rare or new), and L is the sum of all L_i .

- 1.10. Error models (to be discussed what models, BF007) using identical kits
Posterior probability genotype observation correction
- 1.11. Error models (to be discussed what models, BF007) using various kits
Posterior probability mutation correction
2. Profile manipulations
 - 2.1. Combining observations for a sample to “virtual profiles”, for the purpose of collapsing large data sets. (“representative specimen”). This virtual profile is added to the main UI after finding identical DNA profiles.
Many UI-profiles may be collapsed to a virtual single UI-profile. Information about the identical DNA profiles is stored.
 - 2.2. Combining partial results to “consensus-profiles” (BF003)
This is combining partial results / test profiles of one sample into one profile.
3. Static data configuration
 - 3.1. specify DNA-kit parameters per locus and/or group of loci
 - 3.2. functionality for Y-chromosomes
 - 3.3. functionality for SNP-data
 - 3.4. functionality of mtDNA-data
4. Accommodate for Metadata
Import data from PlassData
[NX002] Data encryption (only relevant for non-anonymous metadata)
5. Audit trail
For a project it should be possible to report (Napoleon) an audit-trail (data contained by Bonaparte) showing what happened to it over time.
6. Requirement expansions from phase 1 ([referenced requirement]):
 - 6.1. [NC009] SNP linkage
 - 6.2. [BC002] Storage of peak-heights



Owner : NFI
Project Name : Napoleon
Project Number : 310.158
Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

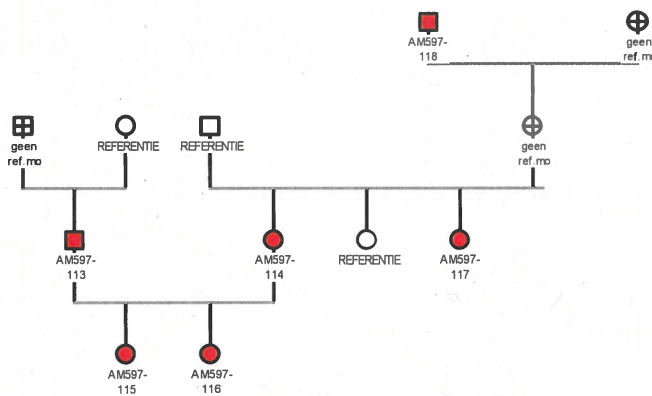
Appendix on pedigrees

This Appendix presents an example Pedigree, used for focusing the discussions.

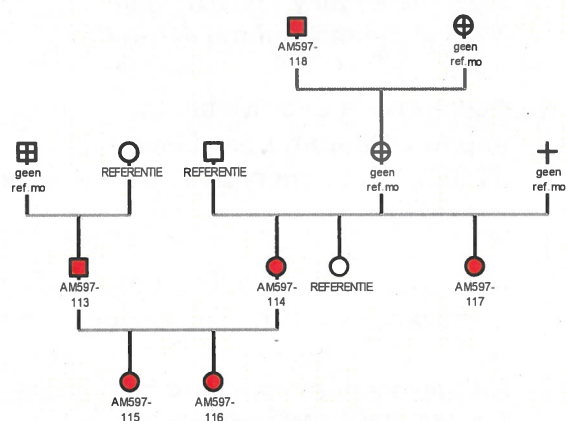
Legend:

- Squares represent males
- Circles represent females
- Red represents a Missing Person
- White represents that there is a reference profile available
- Strikethrough represents that there is no reference profile available
- Children are positioned below and between father and mother (e.g. half-sisters are between the same mother with different fathers)

The example below concerns an putative Pedigree, which had to be adapted during the match-process, because it turned out to be contradicting with 1:1 matches found.



Initial interpretation of Pedigree
(a putative Pedigree)



Eventual interpretation of Pedigree,
after evaluation



Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 4.0
Status : Definition phase 2
Date : 17-11-2009

Appendix on alleles

The basic set of alleles is taken from the OmniPop set⁸.
This set represents all ever observed alleles, independent of the subpopulation.

When we consider a project, three types of alleles can be distinguished per locus:

- Common: allele known on this locus within the subpopulation considered
- Rare: allele known on this locus within the OmniPop set, but not within the subpopulation considered
- New: allele never observed before on this locus within the OmniPop set.

If an UI is compared to a pedigree, all alleles that are in either the UI or the pedigree are recognized as new for that comparison (so only the alleles of the UI and pedigree under consideration are taken into account, not the ones that are in other UIs or pedigrees).

For the size bias correction λ should be dependent on the allele type:

- $\lambda_{\text{common}} = 1$
- $\lambda_{\text{rare}} = 0,5$
- $\lambda_{\text{new}} = 0.1$

Let N_{allele} be the count of the alleles in the subpopulation, so for common alleles $N_{\text{allele}} > 0$, and for rare and new alleles $N_{\text{allele}} = 0$. Furthermore let $\#\text{commons}$ be the total of different common alleles, $\#\text{rares}$ the total of different rare alleles, and $\#\text{news}$ the total of different new alleles. The assumed allele probability is then

$$P_{\text{allele}} = (N_{\text{allele}} + \lambda_{\text{allele}}) / (N + \lambda_{\text{common}} * \#\text{commons} + \lambda_{\text{rare}} * \#\text{rares} + \lambda_{\text{new}} * \#\text{news}).$$

⁸ See e.g.: <http://dnaconsultants.com/Detailed/336.html>

