

## Program of Requirements



## Project Napoleon

**Owner** : Netherlands Forensic Institute  
**Project Name** : Napoleon  
**Project Number** : 310.158  
**Version** : 1.0  
**Status** : Preliminary  
**Date** : 14-07-2008



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

## TABLE OF CONTENTS

<b>1</b>	<b>DOCUMENT INFO .....</b>	<b>3</b>
<b>2</b>	<b>INTRODUCTION .....</b>	<b>5</b>
<b>3</b>	<b>NAPOLEON .....</b>	<b>7</b>
3.1	PURPOSE OF THE NAPOLEON PRODUCT .....	7
3.2	FUNCTIONAL REQUIREMENTS .....	7
3.3	CONSTRAINTS .....	9
3.4	OPERATIONAL REQUIREMENTS .....	10
3.5	RELIABILITY REQUIREMENTS .....	10
3.6	MAINTENANCE REQUIREMENTS .....	10
3.7	SYSTEM REQUIREMENTS .....	11
3.8	SECURITY REQUIREMENTS .....	11
3.9	USABILITY .....	12
<b>4</b>	<b>BONAPARTE .....</b>	<b>13</b>
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 .....	20
<b>5</b>	<b>WORKFLOW .....</b>	<b>22</b>
5.1	CREATING FOLDERS AND PROJECTS .....	22
5.2	DNA-MATCHING .....	22
5.3	MAINTENANCE FUNCTIONS .....	25
<b>6</b>	<b>APPENDIX ON PEDIGREES .....</b>	<b>27</b>



Owner : NFI  
 Project Name : Napoleon  
 Project Number : 310.158

Version : 1.0  
 Status : Preliminary  
 Date : 14-07-2008

# 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	10, 2, e	After discussion with 10, 2, e
0.3	07-06-2008	Concept	10, 2, e	After discussion with SNN on 05-06-2008. Revised by 10, 2, e
0.4	13-06-2008	Concept	10, 2, e	Comments SNN processed.
0.5	23-06-2008	Concept	10, 2, e	Comments of 10, 2, e (ICMP) processed.
0.6	02-07-2008	Concept	10, 2, e	Comments of 10, 2, e processed.
1.0	14-07-2008	Preliminary	10, 2, e	Comments SNN processed. This version is the baseline for phase 1; phase 2 will be elaborated in future versions.

## 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	1.1	

## Acronyms en definitions

Acronym	Description
(CODIS) Case ID	CODIS-key: NFI-case ID + KLPD#
Allele	Number of repeating DNA-blocks on a Locus
AM	Ante Mortem
AM profile	DNA-profile of a relative of a MP or a personal item belonging to a MP
Batch	Search definition within all DNA profiles in a database, based on a selection of characteristics. Batches are not bound to Projects or Folders
BSO	Biologisch Sporen Onderzoek
CODIS	Combined DNA Index System
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 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)
DNA profile	See Profile
DT&B	Digitale Technologie en Biometrie
FM	Family Member (Father, Mother, Child)
Folder	Collection of Projects with a common characteristic, such as "Srebrenica", "Suriname"



**Owner** : NFI  
**Project Name** : Napoleon  
**Project Number** : 310.158  
  
**Version** : 1.0  
**Status** : Preliminary  
**Date** : 14-07-2008

	resp. "DVP" and common match criteria.
GUI	Graphical User Interface
ICIS	Interactive Collaborative Information Systems
ICMP	International Commission on Missing Persons ( <a href="http://www.ic-mp.org">http://www.ic-mp.org</a> )
Individual	Individual person (can be either a UI or a person in a pedigree) An Individual belongs to exactly one Project.
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. A MP can have a Profile.
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
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.
PM	Post Mortem
PM profile	DNA-profile of a victim
Profile	Anonymous DNA profile related to a case. A UI <i>profile</i> is PM, collected from unidentified remains An AM <i>profile</i> is either: <ul style="list-style-type: none"> <li>• collected from a family member of a missing person</li> <li>• collected from a personal item, such as a toothbrush of a missing person</li> </ul>
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"> <li>• 0 or more unidentified individuals</li> <li>• 0 or more (putative) AM profiles. (A putative AM profile can be a PM profile)</li> <li>• 0 or more (putative) pedigrees</li> </ul> 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
SearchSet	Set of UIs to be matched. A SearchSet is part of a Project.
SIN	Specimen Identification Number, represents a SVO
SNN	Stichting Neurale Netwerken
SNP	Single Nucleotide Polymorphism
Specimen Type	A.k.a. 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
SVO	Stuk van Overtuiging
UI	Unidentified Individual or Personal Effect <sup>1</sup>
Virtual profile	Constructed profile, combining two or more partial DNA profiles
XML	Extended Markup Language

<sup>1</sup> 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 : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

## 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 Biologisch Sporen Onderzoek (BSO)
- 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 : 1.0  
Status : Preliminary  
Date : 14-07-2008

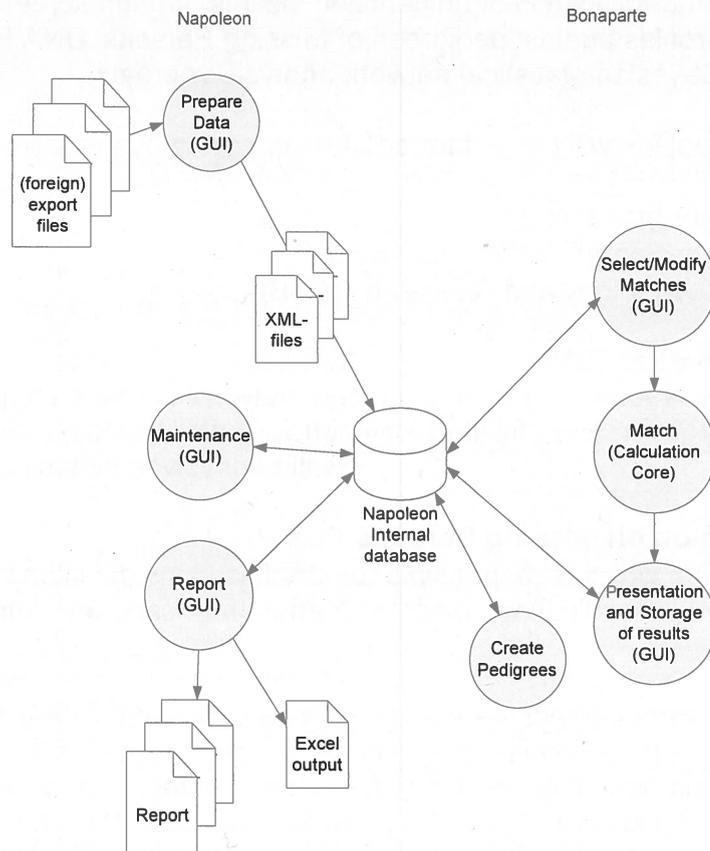


Figure 1: the Napoleon and Bonaparte domains

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 : 1.0  
Status : Preliminary  
Date : 14-07-2008

## 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
- 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 gauge 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 (still to be defined).

Napoleon associates all imported profiles with at least:

- Role
- Project
- Folder

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

Possible Roles:

Role	Type	DNA present	ID known	Description
FM	Reference	✓	✓	Family Member
UI	Target	✓	x	Unidentified Individual
PE	Target	✓	possibly	Personal Effect, either UI (no ID) or FM (ID)
MP	Target	x	✓	Missing Person

#### 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.



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

#### **[NF004] DNA-profile manipulation**

Napoleon shall be able to manipulate DNA profile tables:

- Search
- Update (e.g. add or edit tags)
- Delete
- Add

#### **[NF005] Static Data Configuration**

Napoleon shall support the following Static Data configuration and maintenance options:

- Locus table : adding/update/delete locus
- Amplification Kit table : add amplification kit loci specifics. Data compatibility with all current commercially available kits and markers and extensible for custom systems and marker sets as needed, to ensure long-term usability.
- Population Statistics table: search/add/update/delete population statistics
- Kinship table configuration : add/update/delete kinships
- Mutation Models
- Size bias models
- Test data
- Minimum frequency of rare alleles
- Null alleles
- Dropout alleles (phase 2)

It shall be possible to enter and maintain these sets through a basic GUI.

#### **[NF006] Archiving (phase 2)**

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.

### **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)
4. Deleted-Reports, when a case is (virtually) deleted from Napoleon

#### **[NF008] Report archeology (phase 2)**



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

It may occur that a case has to be re-examined. When that case was already archived (deleted from the current workspace) it should be possible to reconstruct the context of the case report, so that the findings can still be confirmed.

### 3.3 Constraints

#### [NC001]

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

#### [NC002]

The methodological environment of the Napoleon development should be UML, OOD

#### [NC003]

The management environment for the Napoleon project should be Prince2.

#### [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.

In this structure the *Individual* is central.

An Individual (with a DNA-profile) in CODIS is uniquely identified by a CODIS Specimen ID and carry info about the NFI-case ID and the Specimen Type.

#### [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)

#### [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.



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

### 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 made 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:

User Role	Privileges	Comments
Administrator	all	System administrator
Data Enterer	Basic maintenance	Data Entry (profiles, pedigrees, etc.)
Forensic Researcher	Matching, Basic maintenance	e.g. BSO personnel
SNN-user	Maintenance	e.g. for maintenance
Foreign user	Matching	e.g. ICMP-personnel

### 3.5 Reliability requirements

#### [NR001] Backups

There shall be periodic automatic backups so that little 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



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

#### **[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 3<sup>rd</sup> party products**

All 3<sup>rd</sup> 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.

### **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.

Napoleon 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.



**Owner** : NFI  
**Project Name** : Napoleon  
**Project Number** : 310.158  
  
**Version** : 1.0  
**Status** : Preliminary  
**Date** : 14-07-2008

---

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).

**[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 : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

## 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
- Preparing the Match
- Communication Protocol

### 4.2 Functional requirements

#### 4.2.1 Computational Core

The computational core performs the screening and matching. Matching can be done either automatically 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 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)
- kinship match MP vs. MP (not likely)

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 (SearchSet vs. Batch), where x, y are numbers [1,N]

Direct matches:

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



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

#### 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, KIT
  - b. Multiple genotypes with their according probabilities<sup>2</sup>
3. Population allele frequencies
4. Theta-correction
5. Rare alleles
6. Null-alleles
7. Size bias correction models

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 Preparing the Match**

The GUI shall allow the user to

- enter configuration data and selections for the computational core
- perform basic manipulations on profiles, models and sets
- manipulate pedigrees

#### **[BF003]** Profile Management<sup>3</sup>

The following operations are possible with profiles. Reference is the Napoleon internal database. Adapted profiles are not fed back into the source-database.

For profiles constructed through the Napoleon GUI:

---

<sup>2</sup> 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.

<sup>3</sup> The system will compute with at most one (maybe virtual) profile per individual.



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

- 1) Add a profile (NB: an artificial profile must be clearly marked as being artificial)
  - 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) Population<sup>4</sup>
  - b) Pedigree reference(s)
  - c) Alleles
  - d) Add Loci

For external profiles:

- 3) Import profiles from an external source (XML-interface t.b.d.)

For any profile:

- 4) Export profiles to an external file (XML-interface t.b.d.)
- 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

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,
- "Reference AM sample present" = "blank"
- "PE" = green,
- "unidentified MP" = red,
- "identified MP" = blue.

Pedigrees may be composed by dragging Individuals from the Project onto positions in the pedigree. Role information already present in the source data defines a "default" pedigree; this is presented by the Pedigree builder as first guess.

A pedigree (that passes the consistency check, see other requirement) can be saved, printed and stored.

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.

For phase 1 there will only be input/output of pedigrees through the Napoleon GUI.

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.

#### [BF005] Pedigree Consistency Check

During composition the consistency of a pedigree is assessed. Any change in the pedigree (such as by editing it) will be checked in the background. The system will return a consistency measure after each update of the pedigree, and present the number of mismatches contained in it.

<sup>4</sup> 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 : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

Judgment of the overall consistency is left to the user, who may decide to accept and/or store the pedigree, or adapt the pedigree until the pedigree is considered sufficiently consistent.

#### **[BF006] Match Preparation**

The Bonaparte GUI will come up asking which folders and/or projects are going to be processed. The list of new imported data/new projects and earlier defined projects will be presented by the GUI. (Individuals can belong to one or more projects/batches/SearchSets) The user can select from this list.

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 add pedigree(s).

For the match preparation the user should define the SearchSet  
Now the list with Individuals belonging to the selected project is presented in a list, together with the attribute "Specimen category", for the SearchSet.

Next a Batch can be defined within a Folder.  
The following criteria can be used to create a batch:

- Folder Name
- Project name
- Specimen Type

The created batch is the matching set for the SearchSet; it cannot be saved as a separate entity within a Folder, but its use in a Match is logged.

There can be selectable predefined Batches.

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

#### **[BF007] Match Setup (defaults per Folder)**

##### Kinship Matches:

- 1) Select Loci
- 2) Select population for this Match
- 3) Select mutation model or specify a customized mutation model
- 4) Select model for subpopulation correction
- 5) Select model for size bias correction or specify a customized size bias correction model
- 6) Select model for null alleles
- 7) Select model for rare alleles
- 8) Prior odds

##### Direct Matches:

- 1) Select Loci
- 2) Select population for this Match
- 3) Select model for subpopulation correction
- 4) Select model for size bias correction or specify a customized size bias correction model
- 5) Prior odds

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



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

### **[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, SearchSet, Batch, 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, min. # loci, max. # mutations
- Statistics type
  - LR (likelihood Ratio)
  - Posterior odds

Detailed information will be shown on selection of a Hit:

- Date, Issuer, Project, SearchSet, Batch, Match strategy summary, Case ID, Folder
- Pedigree
- LR per Locus
- DNA profiles

The detail report can be exported and printed.

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

Matches must be confirmed by the user. When a UI is confirmed to match with a MP, then the UI is referred to as the Profile of the MP, while the UI refers to the MP for the Identity.

They are not merged within Napoleon, but continue as separate Individuals.

The merging of Individuals as a result of a match 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 can be grouped in Folders.

Both Folders and Projects can be created new.

A Folder may optionally be provided with a Project containing PEs. If so, this Project is subjected to automatic matches performed on the Folder containing it.

### **[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 existing Project
- Put into new Project

The user may now decide to (automatically) update the affected Projects/Folders with a new Search.



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158

Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

#### **[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 contained in so-called Elimination Databases. The content of these databases is clustered in an Elimination Container. This container is like a Project, but without a Folder containing it. Special property of this container is that all new information is automatically matched against it to eliminate Hits caused by contaminations.

#### **4.2.3 Communication Protocol**

Exchange of data between Napoleon and Bonaparte is implemented either:

- by XML-files, containing DNA profiles from source databases plus additional data
- through the common internal Napoleon database (see Figure 1)

Both the content of the XML-files and the data model are still to be elaborated.

Napoleon is responsible for preparing the input data in XML-format, and place it in a workspace. Bonaparte can import this prepared data and export results to the workspace.

#### **4.3 Constraints**

Napoleon's constraints (paragraph 3.3) NC001, NC002, 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 (one genotype with 100% percentage). Peak-heights will not be stored.

#### **4.4 Operational requirements**

Napoleon's operational requirements (paragraph 3.4) NO001 ad NO002 apply to Bonaparte as well.

##### **[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**



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

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).

#### **4.5 Reliability requirements**

##### **[BR001] Search count**

The number of searches performed on a specimen shall be written in the specimen data (like the search count in CODIS DVP).

#### **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 NU001) are automatically applied.

##### **[BU002]**

The statistical model usage shall be transparent to the user. This means that all relevant parameters used in the model shall be presented 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 : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

#### 4.10 Phase 2

This paragraph describes the requirements for the second phase of the project. These requirements will be detailed in future versions of this document.

##### In general:

1. Working with:
  - 1.1. Y-chromosome profiles
  - 1.2. SNP profiles
  - 1.3. mtDNA profiles
2. Construction of consensus profiles: combining test results / test profiles of more than one sample to one profile.
3. Construction of virtual profiles: combining partial results / test profiles of one sample to one profile.
4. Exclusion of paternal or maternal relation in a match

##### For the Calculation core:

5. Expand the Bayesian network with:
  - 5.1. linkage between SNP-markers
  - 5.2. Y-chromosomal data
  - 5.3. mtDNA data
  - 5.4. mutation model of Xu et al, 2000
  - 5.5. genotype observation model with kit-specific parameters (possibly STR peak heights)
  - 5.6. combining observations for a sample (virtual profiles)
6. Expand the Matching with:
  - 6.1. virtual profile calculation
  - 6.2. clustering of profiles into consensus profile
7. QA calculation
  - 7.1. posterior probability genotype error
  - 7.2. posterior probability mutation

##### For the Preparation of Match:

8. Add profile functionality for
  - 8.1. Create profile groups
  - 8.2. Convert profiles (ordered in groups/databases)



**Owner** : NFI  
**Project Name** : Napoleon  
**Project Number** : 310.158  
  
**Version** : 1.0  
**Status** : Preliminary  
**Date** : 14-07-2008

---

- 8.3. Y-chromosomes
- 8.4. SNP-data
- 8.5. mtDNA data
- 8.6. management of virtual profiles
- 8.7. management of consensus profiles
  
- 9. Static data configuration
  - 9.1. specify DNA-kit parameters per locus and/or group of loci
  - 9.2. functionality for Y-chromosomes
  - 9.3. functionality for SNP-data
  - 9.4. functionality of mtDNA-data
  
- 10. Analysis: Construction of virtual profiles
  
- 11. Matching: Construction of consensus profiles
  - Identify/cluster all profiles belonging to one individual



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

## 5 Workflow

This chapter globally (non-SMART) describes the way of working for Napoleon users. Its purpose is to provide context to the Napoleon and Bonaparte functionality, so that the (SMART) engineering requirements may be formulated “fit for purpose”.

The described workflow was defined together with BSO-personnel.

### 5.1 Creating Folders and Projects

Use of Folders and Projects is meant to ease working with large sets of profiles.

A Project is always centered around an Unidentified Individual (UI).

A Folder is a logical set of Projects related by a common characteristic.

A Project may be considered as a workspace containing profiles, pedigrees, configurations, match results, etc. related to matches performed on the Individual.

It can be defined by the user by selecting a profile and giving it a unique name (default name is the SIN). Information can be added to a Project by selecting (e.g. SearchSets), configuring or matching.

A Project can reside in a Folder, bringing together related Projects.

For example, the unidentified profiles of a mass grave in Srebrenica may be clustered in a Folder called “Srebrenica MG”, containing several Projects with UI profiles. Each of these Projects contain the profiles associated with one Individual.

Another Folder might be all the Missing Persons registered in Srebrenica.

Projects and Folders are used in the Bonaparte function “Match Preparation”.

The Search strategy is an (adaptable) property of a Folder.

In phase 2 it will be possible to give a priority to Individuals related to Projects.

### 5.2 DNA-matching

This section describes the workflow for the roles “Forensic researcher”, “Data Enterer” and “Foreign User”.

Matching with Napoleon globally consists of the following steps:

Start up and Login

{

Select or create a Project

{

- Preparation of Match (“what”: select a SearchSet and Batch)



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

- Setting up the Match (“how”: select the parameters and settings for the Match)
  - Match and Analyze Match results
  - (if necessary) Adapt Match configuration
  - Report and/or Store Match results
- }
- Save the Project
- }
- Logout and Close

These steps are described below.

### 5.2.1 Start up

Start the Napoleon application and log in with the proper user role. Napoleon comes up in the same state as this user left it behind after the previous use.

### 5.2.2 Select a Project

The user first defines the environment in which DNA-matching is to be performed: the *Project*. The Project only exists in Napoleon, not within the source DNA-databases.

Projects can be selected from a list to continue previous work with Napoleon, or created new (see earlier). New projects should be given easily recognizable names by the user, for later reference.

Profiles can be added to the Project from the source databases. This is done by selecting Profiles with selection-attributes such as a NFI-casenummer. All Profiles having this attribute are then added to a Project.

A new or adapted Project must be saved before Matches can be performed within it.

### 5.2.3 Match Preparation

#### 5.2.3.1 Create or Select the SearchSet

The SearchSet is the set of target-profiles which is to be matched against a Batch. It can either be created or selected from a list of previously defined SearchSets in the Project.

Creating a SearchSet: select a set of Profiles from the Project by entering a search-category, such as entry-date, gender or NFI-case#. Save this SearchSet under a straightforward name, say “MP”. SearchSets exist of UI-profiles.

Per Match there is always one current SearchSet. A SearchSet belongs to a Project.



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158

Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

### 5.2.3.2 Create or Select the Batch

A Batch is a characteristics-based collection of Profiles. Batches exist outside Folders because they extend over the complete set of currently available Profiles. They can either be created ad-hoc or selected from a list of predefined Batches.

Creating a Batch: select a set of Profiles from the source database(s) by entering one or more search-criteria, such as the entry-date (example: "all Profiles entered in April 2007"). Save this Batch under a straightforward name, say "April 2007".

Batches consist of MP and/or UI profiles and can be created using multiple criteria such as:

- (see [NX002]): Privacy fields
- Project
- Case ID
- KLPD#
- Aanvraag#
- Pedigree
- Profile completeness (number of loci)
- Date
- Specimen Category

Per Match there is always one current Batch. A Batch is independent of the Project.

#### Match Setup

Now that the current SearchSet and the current Batch are identified, the user must define *how* the Match is to be done.

Several parameters and selections can be entered (see requirements) to model the calculations.

Optionally all Match-data can be exported for comparison purposes (outside Napoleon).

The Match is now ready for execution.

### 5.2.4 Match and Analyze Match results

After pressing the "Match" button, Bonaparte will calculate the Match results and present them in a Hitlist with the essential information per Hit: Profile IDs and proof-value.

This provides the user with a summary of the Match results, e.g. :

1 { [tags] PM 597.001 – AM597.001 LR= 10<sup>6</sup> } N

The Hitlist can then be analyzed by the user who can select Hits to see their match-details. Optionally Hit-data can be exported for comparison purposes (outside Napoleon).

Hits may be provided with [profile-tags] to indicate whether a profile "was matched X-times before", is irrelevant, or other (to be determined). Profile-tags can be attached to SearchSet-profiles as well as Batch-profiles.



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

---

The Hitlist may be saved, if considered relevant.

### 5.2.5 Adapt Match configuration

If the Match did not deliver satisfactory results, the user may go back to step Match Preparation, to define another Match and repeat the step Match and Analyze. Default selections are the previously chosen selections.

The new Hitlist will replace the actual Hitlist.

### 5.2.6 Report and Store Match results

Once the user is convinced that a Match is relevant and complete, it can be reported and stored in the Project

When a Match is stored, the following information is saved:

- SearchSet,
- Batch
- Match-configuration,
- Hits
- Pedigrees (if applicable)

To continue Matching the user can now go back to "Select a Project" to switch Project, or "Match Preparation" to continue within the same Project.

### 5.2.7 Save the Project

After the matching work is finished, the Project should be saved. The saved Project contains all SearchSets, Batches and Match results used in the Project so far.

## 5.3 Maintenance functions

These functions are divided into "basic functions" and "advanced functions". This section describes the workflow for the roles "Administrator" (all), "Data Enterer" (basic only) and "Forensic Researcher" (basic only).

To be detailed.

### 5.3.1 Start up

### 5.3.2 System Maintenance

### 5.3.3 User Administration



**Owner** : NFI  
**Project Name** : Napoleon  
**Project Number** : 310.158  
  
**Version** : 1.0  
**Status** : Preliminary  
**Date** : 14-07-2008

---

### **5.3.4 Database Administration**

### **5.3.5 Project Administration**

### **5.3.6 Folder Administration**



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158  
Version : 1.0  
Status : Preliminary  
Date : 14-07-2008

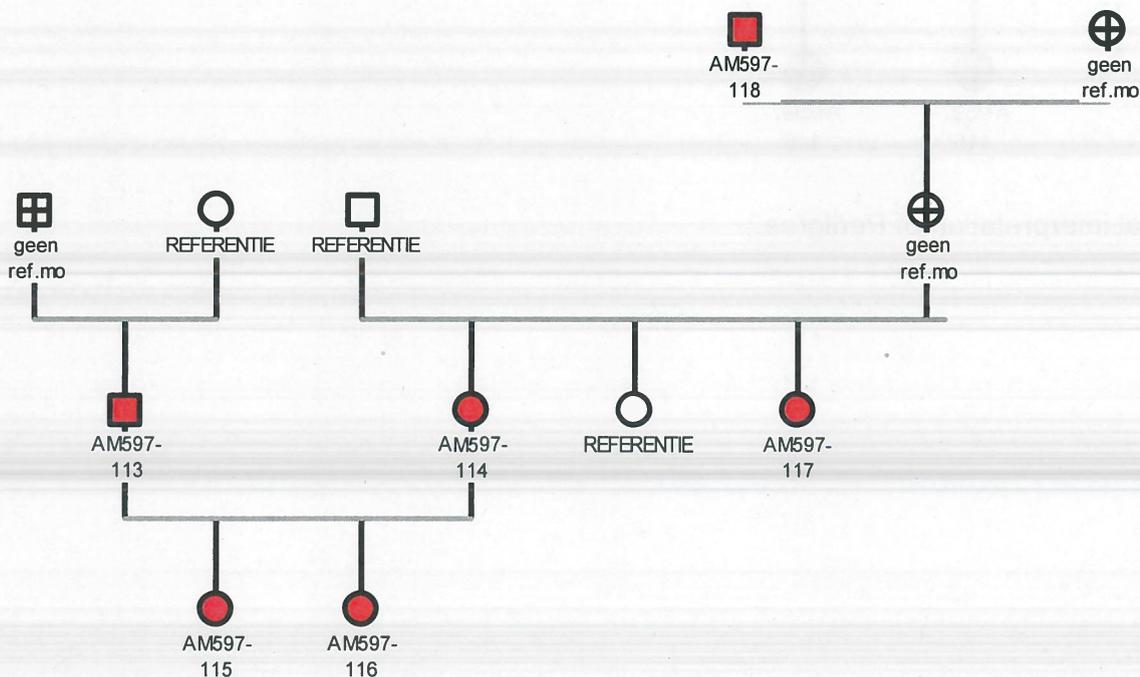
## 6 Appendix on Pedigrees

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

### Legend:

- Squares represent males
- Circles represent females
- Red represents a Missing Person
- Green represents a Personal Effect
- White represents that there is a reference profile available (AM-profile)
- 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.

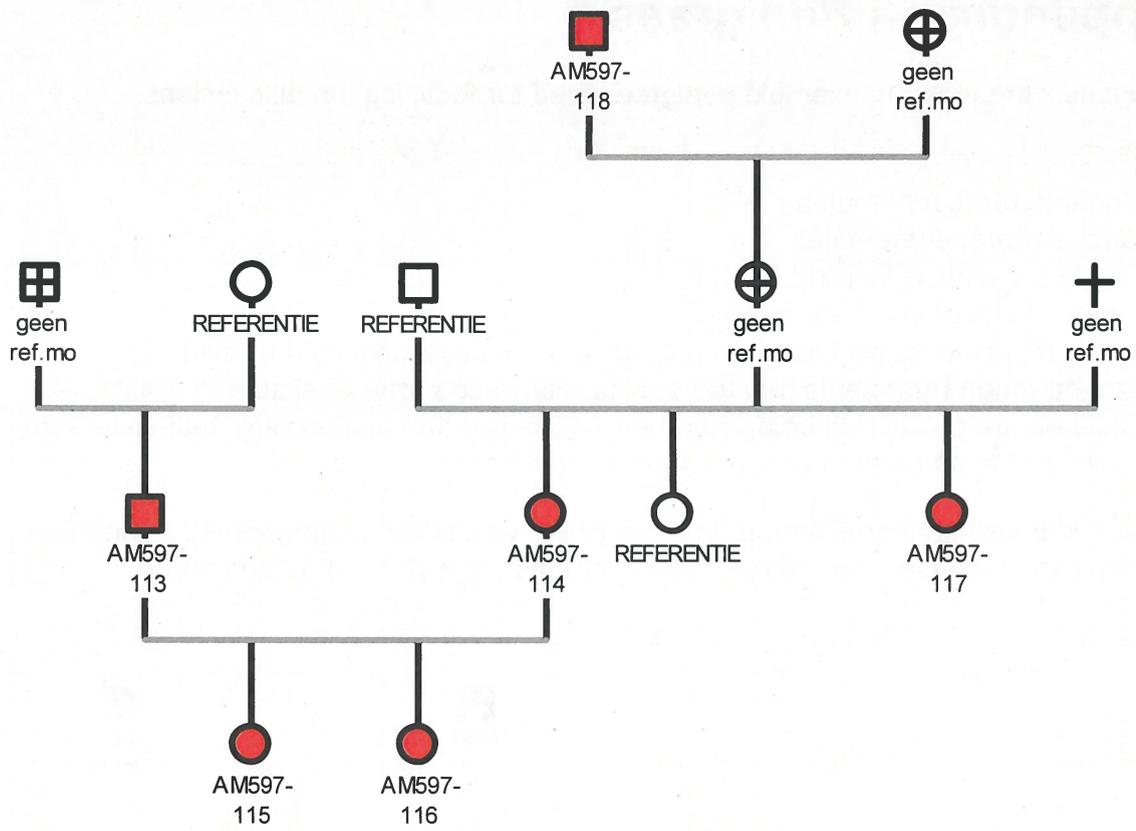


Initial interpretation of Pedigree (a putative pedigree).



Owner : NFI  
Project Name : Napoleon  
Project Number : 310.158

Version : 1.0  
Status : Preliminary  
Date : 14-07-2008



Eventual interpretation of Pedigree.