Programme of Requirements

Project Napoleon

Owner	:	Nederlands Forensisch Instituut
Project Name	:	Napoleon
Project Number	:	310.158
Document Code	:	xxx
Version	:	0.1
Status	:	Concept
Date	:	29-05-2008

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158
Version	: 0.1
Status	: Concept
Date	: 29-05-2008

TABLE OF CONTENTS

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158
Version	: 0.1
Status	: Concept
Date	: 29-05-2008

1 Document Info

Document History

Version	Date	Status	Author	Change description
0.1	29-05-2008	Concept	10, 2, e 10, 2, e	Draft; after discussions with SNN on 22-05-2008 and on 26-05-2008. *Describes requirements for phase 1, with basic functionality.*

Table of References

Reference [ref]	Title	Author	Version	Comments
Napoleon	Research Proposal Napoleon (PID)	NFI	1.1	
Bonaparte	Bonaparte: Bayesian networks for victim identification on the basis of DNA profiles	SNN	2.1	

Acronyms en definitions

Acronym	Description
Allele	Number of repeating DNA-blocks on a Locus
AM	Ante Mortem, data used for identifying missing persons
Batch	Search definition within all DNA-profiles in a database, based on a selection of characteristics, such as gender, finding-place, etc.
BSO	Biologisch Sporen Onderzoek
CODIS	Combined DNA Index System, NFI-database containing DNA-profiles of missing persons and reference persons. (Allele-based and not containing mixed profiles)
GUI	Graphical User Interface
ICMP	International Committee for Missing Persons
Locus	Position of a STR on the DNA
Metaproject	Set of projects with a common factor, such as a time-period or a geographical location.
Mixed profile	DNA-profile of a contaminated specimen, containing DNA-material of more than one individual
MP	Missing Person
NFI	Nederlands Forensisch Instituut
Pedigree	Family of AM-profiles, with their mutual relations
PM	Post Mortem, data used for identifying (remains of) persons
PROMIS	LIMS system of the NFI
SearchSet	Set of profiles to be matched
SIN	Specimen Identification Number, represents a SVO
SNN	Stichting Neurale Netwerken
STR	Short Tandem Repeat, repetitions of tiny pieces of DNA, used for matching because they can vary strongly between individuals
SVO	Stuk van Overtuiging
Virtual profile	Constructed profile, combining two or more partial DNA-profiles
Project	Workset within Napoleon, consisting of one or more DNA-profiles e.g. linked to the same NFI#. A project consists of • 1 or more targets (MP or UI) • 0 or more References (non-target or AM-profiles) • 0 or more Pedigrees
UI	Unidentified person

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158
Version	: 0.1
Status	: Concept
Date	: 29-05-2008

2 Introduction

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158

Version	: 0.1
Status	: Concept
Date	: 29-05-2008

3 Napoleon

3.1 Purpose of the Napoleon product

3.2 Functional requirements

[NF001]
Identification fields used in database structures, such as SIN-numbers, must be flexible so that data from various sources can be processed.
Minimal sources are CODIS and the ICMP-database.

[NF002]
Backups can be made. The backup only concerns the internal Napoleon database.
An automatic backup is made daily.

[NF003]
Napoleon can deliver the following user specified Reports:
   1. …
   2. …
   3.

[NF004]
Napoleon can manipulate databases in the following way:
   - Add a database
   - Delete a database
   - Convert a database with e.g. ETL-scripts (how? which types are to be supported?)
   - Import DNA-profiles from foreign databases (which types are to be supported?)
   - Export DNA-profiles (which formats are to be supported?)

[NF005]
Napoleon shall be able to manipulate DNA-profile tables
   - Search
   - Update
   - Delete
   - Add

[NF006]
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

	Owner	: NFI
	Project Name	: Napoleon
	Project Number	: 310.158
	Version	: 0.1
	Status	: Concept
	Date	: 29-05-2008

3.3 Constraints

[NC001]
The technical environment in which Napoleon is developed should be Oracle/Java and Flexible or Open source platforms.

[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 database structure and datamodel of Napoleon should be aligned with those of the ICMP-system. In a later stage the pedigree module to be developed should be integrable in both systems.

[NC005]
Data-structure for DNA-profiles shall be compatible to the CODIS-datastructure.

In this structure the *specimen* (or Sample) is central.
A specimen is identified by a Specimen ID (KLPD number + SIN number), and carries a CASE ID, which is a unique key for a case in PROMIS (het DNA-zegelnummer?)

[NC006]
Napoleon must be validated with a proper test-set to be defined by NFI.

[NC007]
The interfaces with databases consist of exported files with DNA-profiles in a predefined format.
Napoleon does not export any information into its data-sources

[NC008]
There is no live-interface between Napoleon and any data-providing database..

[NC009]
Linkage (relations between alleles) is not applied within Napoleon.

3.4 Operational requirements

[NO001]
Napoleon will be able to run standalone on a commodity laptop as well as in a network environment with multiple users. The maximum number of concurrent users is 8.

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158
Version	: 0.1
Status	: Concept
Date	: 29-05-2008

[NO002]
All searches and decisions made by the user shall be logged.

[NO003]
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 Roles and according privileges are identified:

Role	Privileges	Comments
Administrator	all	System administrator
Forensic Researcher	Matching, basic maintenance	e.g. BSO personnel
SNN-user	Maintenance	e.g. for maintenance
Foreign user	Matching, basic maintenance within own datasets.	e.g. ICMP-personnel

3.5 Reliability requirements

[NR001]
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.

3.6 Maintenance requirements

[NM001] Compatibility
Napoleon must keep up with:
- new releases of CODIS
- new releases of PROMIS
- new statistical models
- new DNA-kits

[NM002] 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.

[NM003] Documentation
Clear design documentation and manuals are needed for the use and maintenance of Napoleon.

[NM004]
There is a DBA-role for cleaning up the operational environment ("Administrator").

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158
Version	: 0.1
Status	: Concept
Date	: 29-05-2008

3.7 System requirements

[NS001]
- Napoleon will have additional database conversion modules allowing it to work with exports from ICMP and other foreign DNA-databases.

[NS002]
Napoleon shall contain rudimentary static data to be used for Matching:
- Population Sets (some 10 sets of population-related allele frequencies)
- Mutation models (some 5 sets)
- Size bias models
- Test data

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

3.8 Security requirements

[NX001] Authorization
Only authorized users may gain access to Napoleon through a log in procedure.
Napoleon starts up with an authorisation-dialog. After a configurable period of inactivity the authorisation dialog reappears, so that the user has to log in again.

[NX002] Privacy
Privacy-sensitive information shall be encrypted such that it cannot be interpreted without using the Napoleon application.

[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, conform 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	: 0.1
Status	: Concept
Date	: 29-05-2008

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 valorisation proposal [Bonaparte].
This system, the computational core of Napoleon, will be referred to as "Bonaparte".

Bonaparte consists of 3 elements:
- Computational core
- Graphical User Interface (GUI)
- Communication Protocol

4.2 Functional requirements

4.2.1 Computational Core

The computational core performs the matching and shall support the following features:

[BF001] Network builder
Build Bayesian network family – per locus
1. Paternal/Maternal inheritance
   a. Mutation model (such as Ayres and Brenner)
2. Genotype observation
   a. Alleles, KIT
   b. peak heights
3. Population allele frequencies
4. Theta-correction
5. Rare alleles (such as null-alleles)
6. Size bias correction

[BF003] Elementary matching
Elementary matching options:

- x PM – y AM(pedigree) (match the specimen to missing person)
- x PM - y PM (to detect doubles)
- x PM - y PM (to detect parents and siblings)
- x PM - y AM (1:1) to detect parents and siblings
- x PM – y AM(false pedigree)

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158
Version	: 0.1
Status	: Concept
Date	: 29-05-2008

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, models and sets
- manipulate pedigrees

[BF010] Profile Management
The following operations are possible with profiles. Reference is the Napoleon database.
Adapted profiles are not fed back into the source-database.
1) Add a profile
   a) Relatives of MP (AM)
   b) Victim (PM / UI)
   c) Body Part (PM / UI)
2) Edit / Change a profile
   a) Population
   b) Pedigree reference(s)
   c) Alleles
   d) Add Loci
3) Import profiles from an external source (XML-interface t.b.d.)
4) Export profiles to an external file (XML-interface t.b.d.)
5) 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 their status being either
- "no DNA sample" = checked square,
- "Reference AM sample present = "blank"
- "MP" =red,
- "identified PM" = blue.
Pedigrees are limited to 5 generations, and 6 siblings wide.

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.

A pedigree can be saved, printed and stored.

Construction of pedigrees (for 1:1 PM-PM AM-PM matches) pi and si matching should be possible.

[BF005] Pedigree Consistency Check
A (read or constructed through the Bonaparte GUI) pedigree can be parsed for consistency. When an inconsistency is found, the user shall be hinted towards the problem.

Owner	: NFI	
Project Name	: Napoleon	
Project Number	: 310.158	
Version	: 0.1	
Status	: Concept	
Date	: 29-05-2008	

Only pedigrees that pass this consistency check without errors shall be stored and/or used for matching. The Bonaparte GUI shall disable the use and storage of inconsistent pedigrees.

The consistency check is configurable by the user:
− Number of mismatches
− other parameters

The following inconsistencies shall be detected en reported as a result of parsing a pedigree:
1. (grand)child cannot be of a mother / grandparents
2. (grand)child cannot be of a father / grandparents
3. 2 or more siblings can't be full siblings

The user may adapt the pedigree and check again until the pedigree is considered consistent.

[BF006] Match Preparation
The Bonaparte GUI will come up asking which project is going to be processed.
The list of present NFI cases and earlier defined projects will be imported from the DNA-database. (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.

Now the list with specimens belonging to the selected project is presented in a list, together with the attribute Specimen type, for the Search set.

Next a Batch can be selected or created with the GUI.
The following criteria can be used to create a batch:
- Case ID
- KLPD#
- Aanvraag#
- Date (interval) of entry
- Location (remains, or address of MP)
- Complete DNA-profile (Y/N)
- Gender
- Project name
- Pedigree
- Role (in pedigree)
- Age (target)
- Specimen Category
- Ethnicity (standard: NL/Kauk, but can be overruled for modeling)

The selected/created batch is the matching set for the Search Set.
The user may now select the option "match", to start the Match-Setup dialog.

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158
Version	: 0.1
Status	: Concept
Date	: 29-05-2008

[BF007] Match Setup
1) Loci (per profile in SearchSet adaptable)
   a) Add, update, delete locus
2) Select population allele frequencies per profile
3) Select Mutation model (per SearchSet)
4) Select model for subpopulation correction
5) Select model for size bias correction
6) Select model for null alleles
7) Select model for rare alleles
8) Percentages for the probability on a false pedigree (hoort dit niet bij de pedigree zelf?)
9) Thresholds (such as LR, minimum number of loci, max. number of mutations)
10) Statistics type
   a) LR (likelihood Ratio)
   b) Posterior odds
11) Other parameters …..

[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, Search Set, Batch, Match strategy summary, CaseId
2) List of hits 1 { UI/MP | Reference | LR } N

When a hit is selected by double clicking a detail screen pops up, showing:
– Date, Issuer, Project, Search Set, Batch, Match strategy summary, CaseId
– The 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.

[BF009] Store the Match Results
Match results can be stored within the Napoleon database, as part of a Project.
The following data must be saved:

4.3 Constraints

See Constraints in paragraph 3.3.

[BC001]
Mixed profiles are excluded from the computational core.
However the data-storage should be prepared for storing mixed profiles.

	Owner	: NFI
	Project Name	: Napoleon
	Project Number	: 310.158
	Version	: 0.1
	Status	: Concept
	Date	: 29-05-2008

[BC002]
The Bonaparte database shall store percentages (DNA peakheight) instead of interpretations.

[BC003]
The interface between Napoleon and Bonaparte (communication protocol) shall be specified and implemented in XML.

4.4 Operational requirements

4.5 Reliability requirements

[BR001]
The number of false hits per match shall not exceed 0.01%

[BR002]
The number of searches performed on a specimen shall be written in the specimen-data (like the search-count in CODIS).

4.6 Maintenance requirements

4.7 System Requirements

4.8 Security requirements

Napoleon's security requirements (paragraph 3.8), as far as applicable, apply to Bonaparte as well.

4.9 Usability

[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

[BU003]
A user training shall be part of the delivery.

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158
Version	: 0.1
Status	: Concept
Date	: 29-05-2008

5 Workflow for the Forensic Researcher

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

The described workflow was defined together with BSO-personnel.

5.1 DNA-matching

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)
>> - 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.

5.1.1 Start up

Start the Napoleon application and log In as a Forensic Researcher.
Napoleon comes up in the same state as this user left it behind after the previous use.
....

5.1.2 Select or create 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. New projects should be given easily recognizable names by the user, for later reference.

		Owner	: NFI
		Project Name	: Napoleon
		Project Number	: 310.158
		Version	: 0.1
		Status	: Concept
		Date	: 29-05-2008

Profiles can be added to the Project from the source databases. This is done by selecting Profiles with selection-attributes such as a NFI-casenumber. 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.

5.1.3 Match Preparation

5.1.3.1 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.

Creating a SearchSet: select a set of Profiles from the Project by entering a search-category, such as the Specimen Type (example: all Missing Persons).
Save this SearchSet under a straightforward name, say "MP".

SearchSets are of one of the following categories:
1. AM (ante mortem reference profiles)
2. PM (post mortem victim profiles, set 1)
3. Pedigree (set of AM kin-profiles)

Per Match there is always <u>one</u> current SearchSet. A SearchSet belongs to the Project.

5.1.3.2 Create or Select the Batch

The Batch is a characteristics-based collection of Profiles. Batches exist outside Projects, because they extend over the complete set of currently available Profiles.
It can either be created or selected from a list of previously defined 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 can be created using multiple criteria such as:
- Project
- Pedigree
- Role in Pedigree: O (unidentified), F (father), M (mother), A (personal item)
- Gender
- Specimen Category

Per Match there is always <u>one</u> 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.

	Owner	: NFI
	Project Name	: Napoleon
	Project Number	: 310.158
	Version	: 0.1
	Status	: Concept
	Date	: 29-05-2008

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.

5.1.4 Match and Analyze Match results

After pressing the "Match" button, Bonaparte will calculate the Match results and present them in a Hit-list with only the essential information per Hit: Profile IDs and proof-value. This provides the user with a summary of the Match results.

The Hit-list can then be analyzed by the user.
The user may double-click on a Hit to see the match-details (or press "Details")

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

Hits may be provided with tags to indicate whether it "was matched X-times before", is irrelevant, or other (to be determined). The Hit list may be saved, if relevant.

5.1.5 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 Hit-list will replace the actual hit-list.

5.1.6 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.

5.1.7 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.

Owner : NFI
Project Name : Napoleon
Project Number : 310.158

Version : 0.1
Status : Concept
Date : 29-05-2008

5.2 Maintenance functions

To be detailed.

5.2.1 Start up

Start the Napoleon application and Log In as Administrator.
….

5.2.2 System Maintenance

5.2.3 User Administration

5.2.4 Database Administration

5.2.4.1 Profiles

5.2.4.2 Batches

5.2.5 Project Administration

Owner	: NFI
Project Name	: Napoleon
Project Number	: 310.158
Version	: 0.1
Status	: Concept
Date	: 29-05-2008

6 Vragen

1. Moet Napoleon standalone kunnen draaien?

- Voordelen: je kunt een "veldversie" meenemen naar een locatie waar geen verbindingen zijn. (alleen een voordeel op plekken waar geen open internet beschikbaar is)
- Nadelen:
  o bij terugkomst zal de Napoleon database gesynchroniseerd moeten worden.
  o er zijn dan mogelijk twee versies te onderhouden.

2. Kan Napoleon remote access bieden?

- Voordeel: logistiek voordeel bij onderhoud (kosten!)

Bij DT wordt gebruik gemaakt van zgn. DMZs. Dit zijn afgeschermde subnetwerken die benaderbaar zijn vanaf internet, maar waarvandaan niet op interne netwerken kan worden gewerkt (door gebruik van firewalls). Andersom zijn de DMZs weer wel toegankelijk vanaf de interne netwerken.
Zo'n DMZ leent zich dus prima voor remote access: omdat Napoleon werkt met exportbestanden en geen live-koppelingen met databases, kan de complete Napoleon context in een DMZ worden opgenomen. Van buitenaf kan dan zowel de gebruiker als onderhoud toegang krijgen tot Napoleon.

3. Data integriteit?

Binnen Napoleon wordt informatie gecreëerd en opgeslagen die niet in de broninformatie zit, en daar ook niet terechtkomt (er is immers geen koppeling).
Wat is de status van deze informatie? Wat dient ermee te gebeuren als een case wordt gesloten, bijv.?