

**Requirements Specification**

**For**

**DNA Data Extraction**



**Prepared by:**

**Group Name:**

|  |  |
| --- | --- |
| Ben Sanders | sanders.199@wright.edu |
| Jordan Ellis | ellis.145@wright.edu |
| Caleb Smith | smith.2026@wright.edu |
| Frank Randazzo | randazzofr@gmail.com |
| April Stewart | stewart.300@wright.edu |

**For:**

**Project Sponsor: *Wright State University***

|  |  |
| --- | --- |
| Nathaniel Adams | adams.201@wright.edu |

**Document Tracking Information:**

**Document Version *1.0***

**Date**

**Final Approval Date *<XX/XX/XX>***

**Final Approval By: *<name> <e-mail>***

Contents

[**1**](#_30j0zll) **Introduction 4**

[1.1](#_3znysh7) Document Purpose 4

[1.2](#_tyjcwt) Product Scope 4

[1.3](#_4d34og8) Intended Audience and Document Overview 4

[1.4](#_17dp8vu) Definitions, Acronymns, and Abbreviations 5

[*1.4.1*](#_3rdcrjn) *TBD 5*

[1.5](#_lnxbz9) References and Acknowledgments 5

[*1.5.1*](#_35nkun2) *ABIF file format specification: 5*

[*1.5.2*](#_1ksv4uv) *NCBI BatchExtract 5*

[*1.5.3*](#_44sinio) *IEEE Standards 5*

[**2**](#_z337ya) **Overall Description 6**

[2.1](#_1y810tw) Product Perspective 6

[2.2](#_2xcytpi) Product Functionality 6

[*2.2.1*](#_3whwml4) *Compatibility 6*

[*2.2.2*](#_1pxezwc) *Output 6*

[2.3](#_3o7alnk) Design and Implementation Constraints 7

[*2.3.1*](#_23ckvvd) *Hardware 7*

[*2.3.2*](#_ihv636) *Operating System 7*

[2.4](#_1hmsyys) User Documentation 7

[*2.4.1*](#_41mghml) *User manual 7*

[2.5](#_vx1227) Assumptions and Dependencies 7

[**3**](#_1v1yuxt) **Specific Requirements 8**

[3.1](#_2u6wntf) External Interfaces 8

[*3.1.1*](#_3tbugp1) *Command line interface 8*

[*3.1.2*](#_nmf14n) *Hardware Interfaces 8*

[*3.1.3*](#_1mrcu09) *Software Interfaces 8*

[*3.1.4*](#_2lwamvv) *Communications Interfaces 8*

[3.2](#_3l18frh) Functional Requirements 8

[*3.2.1*](#_206ipza) *Input 8*

[*3.2.2*](#_4k668n3) *Schema 9*

[*3.2.3*](#_1egqt2p) *Output 10*

[**4**](#_3ygebqi) **Other Non-Functional Requirements 11**

[4.1](#_sqyw64) Performance Requirements 11

[4.2](#_1rvwp1q) Safety Requirements 11

[*4.2.1*](#_4bvk7pj) *Data integrity 11*

[*4.2.2*](#_2r0uhxc) *Unspecified data tags 11*

[4.3](#_1664s55) Security Requirements 11

[4.4](#_3q5sasy) Software Quality Requirements 11

[4.5](#_25b2l0r) Hardware Quality Requirements 11

[**5**](#_kgcv8k) **Other Requirements 12**

# Introduction

## Document Purpose

This document is a requirement specification for DNA Data Extraction. The specified product is a utility for extracting data from binary .fsa and .hid files generated during the course of DNA testing. The specified product offers greater accessibility to these file contents by extracting them and preserving them in a human- and machine-readable format.

## Product Scope

The overall goal of this project is to construct a data extraction utility that includes the following capabilities:

1. Compatibility with .fsa and .hid files generated by the families of ABI 310, 31xx, and 35xx genetic analyzers and their according data collection programs.
2. Extracts raw, binary signal data from these files.
3. Extracts metadata from these files.
4. Preserves this data/metadata in a human- and machine-readable format.
5. Operated from a command line interface.

Delivery of a product that implements the above capabilities as described in this and other documents shall constitute a complete deliverable.

## Intended Audience and Document Overview

The intended audience of this document is the project sponsor(s) and designers/developers.

Users may include end-users, system administrators, and future developers.

## Definitions, Acronymns, and Abbreviations

### TBD

## References and Acknowledgments

### ABIF file format specification:

#### <http://www6.appliedbiosystems.com/support/software_community/ABIF_File_Format.pdf>

### NCBI BatchExtract

#### National Center for Biotechnology Information (NCBI). BatchExtract. 2006. See website <ftp://ftp.ncbi.nlm.nih.gov/pub/forensics/> (dead link)

### IEEE Standards

#### IEEE Std 830-1998

The IEEE Recommended Practice for Software Requirements Specification

# Overall Description

## Product Perspective

This project intends to deliver DNA Data Extraction, a standalone application that can be utilized to extract and preserve data as described above.

## Product Functionality

### Compatibility

#### Genetic analyzers

DNA Data Extraction shall be compatible with files generated by the families of ABI 310, 31xx, and 35xx genetic analyzers.

#### Data Collection Software

DNA Data Extraction shall be compatible with files generated by all versions of ABI Data Collection Software published in 2016 and earlier.

#### File formats

DNA Data Extraction shall be compatible with both .fsa and .hid file types.

### Output

#### Contents

DNA Data Extraction shall translate and preserve all data specified by tags described in the ABIF file format specification.

#### File type

DNA Data Extraction shall preserve all output data in a human- and machine-readable format.

#### Schema

DNA Data Extraction shall adhere to an internal schema for data preservation/output. Operating Environment

## Design and Implementation Constraints

### Hardware

DNA Data Extraction shall be agnostic to hardware.

### Operating System

DNA Data Extraction shall operate with full functionality on Windows, Apple OSX, and Linux.

Linux testing shall be performed on Ubuntu.

## User Documentation

### User manual

A user manual shall be provided with each distributed copy of DNA Data Extraction or made available on a public-facing website linked to by the program.

## Assumptions and Dependencies

TBD

# Specific Requirements

## External Interfaces

### Command line interface

#### Input

The command line interface shall accept a file or folder as an input parameter.

#### Output

The command line interface shall accept a file or folder as an output parameter.

#### Recursion

The command line interface shall accept a specifier for recursion.

#### User interaction

The command line interface shall require no more than one human interaction with the program via the command line interface. All parameters shall be provided to the program at once.

### Hardware Interfaces

There are no requirements for hardware interfaces.

### Software Interfaces

There are no requirements for software interfaces.

### Communications Interfaces

There are no requirements for communication interfaces.

## Functional Requirements

### Input

#### Input file state

Aside from locking input files, DNA Data Extraction shall not affect the state or content of any file provided as input.

#### Parameters

DNA Data Extraction shall accept as input a path pointing to one file OR one folder.

##### Input

###### Input File

The input file shall be checked for a .fsa or .hid extension. The user shall be notified if the file is incompatible.

###### Input Folder

The contents of the folder shall be checked for .fsa or .hid extensions. Only .fsa and .hid files shall be operated on.

##### Output

###### Output file

If an input file is specified, the output shall be written to a file at the provided output path. If a file of the specified name exists at that path, that file shall be overwritten with no warning to the user.

###### Output folder

If an input folder is specified, the output shall be written in the same relative subdirectories with the same file names (except for file extension) as included in the input folder.

##### Recursion

If a recursion specifier is provided, DNA Data Extraction shall recurse through the input directory’s subdirectories. If no recursion specifier is provided, DNA Data Extraction shall operate on only those files contained directly within the input directory.

### Schema

A hierarchical schema shall be specifiable as a program configuration.

#### Format

The schema shall be specified in a human-readable format.

#### Key-value

The schema shall be specified in a key-value relationship where keys are the data tags specified in the ABIF file format specification and the values are a structure of number, type, and description.

#### Hierarchy

The schema shall be specifiable in a list. E.g. one key may point to multiple subkeys in a list.

#### Multiplicity

A data tag may exist only once within the schema, except under conditions of compatibility.

#### Compatibility

If necessary, tags that exist in multiple genetic analyzer or Data Collection Software file format specifications with different specifications shall be differentiated in the schema by machine model and/or Data Collection Software versions.

#### Unspecified tags

Custom data tags not specified in the ABIF file format specification shall be supported by the schema.

### Output

#### Format

The output format shall be human- and machine-readable. (E.g. XML, JSON, YAML, etc.)

#### Order

The order of tags in the output file shall match the order and hierarchy of the tags in the configuration schema.

#### Scope

All data in an input file matching a data tag specified in the ABIF file format specification shall be preserved in the output file.

# Other Non-Functional Requirements

## Performance Requirements

There are no performance requirements.

## Safety Requirements

### Data integrity

All data for data tags specified in the ABIF file format specification shall be translated with full fidelity to the output file.

### Unspecified data tags

Data tags not specified in the ABIF file format specification have no guarantee to reliability or fidelity of translation.

## Security Requirements

There are no security requirements.

## Software Quality Requirements

There are no software quality requirements.

## Hardware Quality Requirements

There are no hardware quality requirements.

# Other Requirements

There are no other requirements.