



XANALYS PowerCase

White Paper

Copyright and Trademarks

XANALYS PowerCase White Paper

Version 2.1

August 2004

Part number 0603-WP-PWC01 (last revised February 2009)

Copyright © 2009 by Xanalis Limited.

Companies, names and data used in examples herein are fictitious unless otherwise noted.

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of Xanalis Corporation.

The information in this publication is provided for information only and is subject to change without notice. Xanalis Corporation and its affiliates assume no responsibility or liability for any loss or damage that may arise from the use of any information in this publication. The software described in this book is furnished under license and may only be used or copied in accordance with the terms of that license.

Xanalis, the Xanalis logo, PowerCase, Quenza, PowerIndexer and Watson are registered trademarks of Xanalis Limited.

Other brand or product names are the registered trademarks or trademarks of their respective holders.

Xanalis Limited

Market Court
20-24 Church Street
Altrincham, Cheshire
WA14 4DW
United Kingdom

telephone +44 161 941 7792
fax +44 161 332 8285

www.xanalis.com

Table of Contents

1 Introduction..... 1

2 What is XANALYS PowerCase? 1

3 The XANALYS PowerCase Application Framework – An Overview..... 1

4 What Does XANALYS PowerCase Do?..... 2

 Incident Room Procedures and XANALYS PowerCase Support 3

 Data Entry 4

 Indexing & PowerIndexing 5

 Search, Research and Analysis..... 7

 Document Workflow and User Authorization 9

 Case Linking and Triggering. 10

 New Features – PowerCase 4.4 10

1 Introduction

This document is intended to provide an overview of XANALYS PowerCase. It describes what XANALYS PowerCase is, what it does, and gives some indication of the technologies that underpin it.

2 What is XANALYS PowerCase?

XANALYS PowerCase is an Investigative Management System designed to support the common workflow processes identified as hallmarks of best practice in the maintenance and management of evidence during the course of investigation. In this sense Investigative Management concerns the management of all processes involved up to and including the presentation of material for principled scrutiny, for example, the criminal justice system be it in the private or public sector. Most importantly it covers two major processes needed to ensure a successful outcome:

1. Internal Process - The management and processing of all documents, such as witness statements, reports, messages and such like, which are obtained during the course of an investigation.
2. External Process - The management of actions to be undertaken or considered. Actions are essentially leads or lines of enquiry which could or should be pursued.

A third process focuses on the intelligence-led nature of the investigation.

3. The construction and maintenance of a structured entity-relationship database model of people, locations, vehicles, property, events and more importantly the inter-relationships between such entities. This entity-relationship model forms the 'intelligence' gathered on an investigation, and is particularly important in establishing the direction that an enquiry should take.

An investigative management system plays an important role during all stages in the pursuance of a case, from the early investigative stages, to when the case is taken to court, to reviewing and auditing an investigation.

3 The XANALYS PowerCase Application Framework – An Overview

In keeping with the approach adopted for XANALYS Link Explorer (*formerly Watson*), XANALYS PowerCase is built using an Application Framework which employs a collection of definitional tables to describe both the database schema and the form designs to be employed within an "Instantiation" of the FrameWork. An instantiation is a specific application for a specific domain built using the framework. The intention is that this application framework can be employed for a whole class of applications, which can be characterized by combinations the following requirements:

- A basic need to extract information from textual documents.
- A need to organize this extracted information as a structured, entity relationship model.
- A need to undertake various forms of analysis on the entity relationship model so constructed.

- A need to carefully maintain the provenance chain from analytical results through to original source material.
- A need to manage and control the processing of documents.
- A need to raise, prioritize and manage actions to pursue various lines of enquiry.

Clearly the Law Enforcement domain falls into this class of application. Other examples might be

- In litigation support where information from large numbers of documents needs to be exploited, usually in short periods of time.
- In processing the results of drug trials, both in terms of the efficacy and the side effects.
- In tracking and analyzing maintenance and safety reports in complex, expensive or safety critical machinery.
- In research in general where many sources of information have to be reviewed and somehow organized to extract and promote new knowledge.
- In software development where requirements have to be analyzed and organized into some sort of design.

In terms of XANALYS PowerCase as an Investigative Management System in the Law Enforcement market, there are currently instantiations for the UK, US, and Australian markets, plus ones for specific customers. For ease of maintenance these instantiations share a common database schema. However each instantiation differs in terms of form design, labels and names.

4 What Does XANALYS PowerCase Do?

XANALYS PowerCase supports a defined set of procedures that should operate in an “incident room” or “case office.” These procedures were ascertained through a detailed requirements capture process, and have been chiefly influenced by Home Office Procedures in the UK. Although these procedures originate in the UK they are widely accepted in the law enforcement community as being representative of best practice. It is interesting to note that XANALYS PowerCase as a product is perhaps as much a ready made Methodology as it is a Software Application.

Incident Room Procedures and XANALYS PowerCase Support

To understand what XANALYS PowerCase actually does it is useful to first understand some aspects of these procedures. This section gives an overview.

The primary sources of information to an investigation are witness statements, officer reports and incoming telephone messages. In what follows the processing of such a document will be described in some detail. The processing of other kinds of documents all follow a very similar structure.

1. The document is “registered” on the system. This registration process is primarily to record the existence of a document, and as such makes the entire investigation team aware of its availability. Registration also associates, or links, the document with the record of the person who has given the statement, written the report, or telephoned with a message. In other words each document is linked to its source. This is an important principle in XANALYS PowerCase, which by design maintains provenance (origin) links for all material.

XANALYS PowerCase gives assistance in researching for the correct person record (see research section later) and then in automatically creating appropriate links.

2. The details of the document (e.g. name and address), together with the text, are entered into the system. Through using the application framework it was possible to tailor the forms design in XANALYS PowerCase to match printed stationery, and thus facilitate data entry. XANALYS PowerCase also incorporates standard word processing facilities to support the entry of textual material. The content of the document and associated details are then “indexed.” This essentially means extracting information from the text and fields of the document, collating this information as preparatory entity or link records, and then researching the database for similar records. Depending on the result of the research step, either new records will be created or else any new information will be consolidated into existing ones. Either way, links between extracted entities and the source document will be automatically created to provide a complete “provenance” chain between all information recorded in the database and its source. Specifically it is possible to ask detailed questions such as “Who says that Bill Clinton is an associate of Oliver North?” and be directed to each and every document which supports such a link.
3. Actions arising from the document content are then considered. For example an action might be to take a statement from a person mentioned in the text of a statement, or to trace the owner of a vehicle. As with indexing entities, any new action is first researched to ensure that there are no duplicates. When created they are linked to the source document and the “object” of the action, e.g. the person from whom the statement is to be taken, or the vehicle to be traced. Actions form an important in the overall information collection cycle. In one sense actions should be raised for every conceivable line of enquiry, thus leaving “no stone un-turned”. In practice the nature and number of actions raised is governed by the overall strategy or policy set by the senior investigating officer.
4. Actions are allocated to field or enquiry officers to actually undertake the particular line of enquiry captured in each action. All actions are prioritized upon creation, and these priorities can influence their allocation. Typically actions are allocated on a “theme” basis, for example relating to the scene of the crime, such that the same investigation team will deal with all related actions. Some actions will be deferred or even filed according to their relative priority and the overall policy set by the senior investigating officer.

5. On completion, the results of an action are recorded on the system. Any new documents are first registered and then linked to the generating action, thus preserving the provenance of all information. Once this has been completed the action can be closed and any resulting documents processed independently in the same manner described above.

The entire process is quite a simple cycle: register a document, enter its details, index it, raise actions arising from it, the results of actions give rise to new documents and so the process continues. It is clear that following such a process in too simplistic a fashion would give rise to an almost combinatorial explosion of documents, actions to perform, and information to analyze. This is quite a common phenomenon, and care must be taken not to substitute “blind data gathering” for “true intelligence gathering”. XANALYS PowerCase provides support in a number of ways to assist the user with this goal:

- Document and work-flow management: during the document registration process, XANALYS PowerCase encourages the user to enter details about what tasks need to be performed on the document, and also what priority should be assigned to these tasks. From then on XANALYS PowerCase manages the flow of this document through the system, prompting the user to enter the status of the document after each operation, and presenting new tasks to users as the processing of the document progresses. XANALYS PowerCase also provides a number of management review tools in order to pin-point any bottlenecks or problem areas.
- XANALYS PowerCase maintains a provenance chain for every piece of information entered into the system. Each document is registered to an author; each document is linked to the action which gave rise to it; each entity or link is linked to the document or document(s) from whence it came; each attribute value is ultimately linked back to a line or lines of text of a document or documents from whence it was derived. This may seem a little extreme, but in general information or intelligence is of little value unless one also knows how reliable that information is, and this reliability is most often dependent on its source and number of sources.
- XANALYS PowerCase provides various Action Management facilities for prioritizing and organizing actions into themes or queues for later allocation to enquiry officers. This is an absolutely essential requirement, and one which ultimately governs the direction of the entire investigation.
- Finally, through a tight integration with its companion product, XANALYS Link Explorer, XANALYS PowerCase provides the means to analyze the resulting information as it is gathered. XANALYS Link Explorer is a sophisticated data visualization and analysis tool, which can assist the user in the construction of complex queries without database of SQL knowledge, and then visualize the results through a variety of graphical chart layouts.

Data Entry

XANALYS PowerCase supports document update forms, which have been designed to match the printed stationery, thus facilitating data entry. Like many forms based systems, it supports the usual validation rules, drop down lists, cut and paste and other similar facilities for data entry. Each form is thus the electronic equivalent of the paper version, with the implication that once a document has been entered there should seldom be a need to refer to the original manuscript

version. Indeed one of the design goals behind XANALYS PowerCase was the concept of an electronic incident room.

An important design feature is that entities and relationships are created through the process of indexing documents (see later), and not through direct update forms, (although this is permitted in exceptional circumstances). The main reason for this is the desire to maintain a provenance chain or audit trail of the source of each piece of information. If the user was encouraged to edit records directly, then all that could be recorded would be the name of the user making the edit. In XANALYS PowerCase any piece of information can be traced back to a specific line of text in one or more documents.

XANALYS PowerCase also supports multi-media in two important ways:

- First allows for alternative representations of the same document. For example, initially a scanned version of the paper document can be attached to the XANALYS PowerCase document such that investigators can at least read it on line. Later the document may either be typed or OCR'ed in preparation for the indexing process.
- Second various supplementary information such as, for example a photograph, spreadsheet, or a video clip, can be attached to a XANALYS PowerCase document. These attachments can then be "launched" and viewed using some 3rd party software.

Indexing & PowerIndexing

XANALYS PowerCase supports both the manual and automatic indexing of documents through an intermediate presentation structure called the Index Plan. Indexing is defined as the process of extracting information from either the text of a document or associated structured details, and constructing an entity relationship model of the information content. For example the structured fields of the Message form will capture the surname, forenames and address of the person from whom a message originates. The indexing process will automatically extract all information present within these structured details into the index plan. The textual part of the message may, for example, describe a number of visits received and the vehicles they arrived in. In the indexing process all these entities and relationships would somehow have to end up in the structured database. The Index Plan is designed to provide a work around in which such information can be collated and organized in preparation for its eventual transfer into database records.

Each document has an index plan which is displayed on the right hand side of the document form. The Index Plan is a folding outline structure which is used to prepare and collate the information which is to be extracted from the entire document. Its structure is hierarchic, with its root, level 0 node, effectively representing the document as a whole. Level 1 nodes represent Entities or Links between entities which are to be extracted. Level 2 nodes hold Entity Attributes and their proposed Values. Level 3 are Text Extracts representing the textual source of an attribute value. Text Extracts are in fact hyper-linked to the precise location in the text of a document in which they occurred. To continue with the example of the message, Level 1 entity nodes will be prepared for the originating person and that person's address. The visitors and their vehicles will also appear as Level 1 entities. In addition a number of events will be added representing the visits. Each of the events will include implied links to each of the participating entities, namely the people involved and also the location and vehicles. The Level 2 nodes will be populated with Attribute-Value pairs such as Surname/Smith, Forename1/Mick for the originating person, and StreetName/Princes Drive, Town/Marple for the address, and so on.

XANALYS PowerCase will automatically prepare such index plan nodes from the details entered on the structured fields of a document form. Once prepared the user is free to edit these, and add new entities or attributes as appropriate. It also supports both manual Indexing and PowerIndexing of the text of a document.

Manual Indexing

If indexing is performed manually the user will read the text of the document on the left hand side of the screen, and edit the index plan on the right. Upon encountering, for example, a person's name in the text, a new level 1 node can be added to the index plan, via a simple menu command. After this the user can swipe text in the text on the left and use the "Extract As" command to add the swiped text as the value of an attribute in the index plan. A hyper-link is constructed between the node created in the index plan and the place in the text where the value was mentioned. If a value is referenced in the text more than once, then several such hyper-links can be constructed to each and every reference. This hyper-linking is quite valuable when, for example, reviewing a document with respect to what it says about a particular individual.

When the user is satisfied with what has been prepared in the index plan the next step is to research the database in order to check whether such a record already exists. The new information will either be saved into a new record or else consolidated into an existing one depending on the results of the research. The research process will be described in a later section, but here it is worth mentioning some methodological issues which underpin the index plan.

- First and foremost it is a working area in which to collate and organize the information which is to be extracted from the document. In this sense it is a distillation or summary of the content of the document.
- Second, once all research is complete it is a representation of what was indexed from that document.
- Third it is part of the provenance chain maintained by XANALYS PowerCase. Every entity record will have a link to the document or documents from which its attribute were extracted. The index plan takes this provenance linking that one step further in that the user, through the hyper-linking, and return all the way to the exact line of source text from which the attribute was derived.

Automated Indexing or PowerIndexing

Manual indexing is extremely labour intensive. For this reason XANALYS PowerCase includes a natural language processing facility called PowerIndexing. This facility parses the text of a document, driven by a set of grammar rules and a lexicon, marking up the text and extracting entities and their attribute values as it goes. The end result is an automatically prepared index plan, which can be reviewed or directly researched by the end user. Given that manual indexing is such a labour intensive process, PowerIndexing is an extremely popular feature of the XANALYS PowerCase product.

PowerIndexing can be customized for the different variants of English (British, American, Australian etc.), and it can also be customized to particular domains. It is driven by a number of data files:

- **A grammar:** This is represented by a set of grammar rules which essentially define the structure of the phrases or sentences that are to be recognized. The grammar is expressed in a BNF like language, augmented with directives for the information extraction component of the PowerIndexing.
- **A Lexicon:** This is essentially a collection of word lists, which combined with the grammar rules determine what can be recognized.
- **A Custom File:** Elements of the Grammar and Lexicon can be enabled/disabled through directives held in the Custom File. In this way both national differences and specific customer differences can be achieved through simply editing a text file.

XANALYS PowerCase also implements a feed back loop, or learning capability into the operation of PowerIndexing. For example as a case database develops so too does, for example, the list of known surnames. This feature in XANALYS PowerCase allows such known values to be automatically fed into the natural language recognition process, and in so doing improve the recognition accuracy.

Search, Research and Analysis

XANALYS PowerCase combines the power of three complementary technologies in a single application, namely:

- **Research:** The goal here is to determine whether an entity is already present in the structured entity relationship database. It is by nature a structured search, and one which is required to be quite comprehensive since it is important to avoid the creation of duplicate records.
- **Free Text Search:** this is a broadly applicable and rapid technique, used for finding relevant material for subsequent, more detailed examination and often before indexing has been undertaken. Alternatively it can be used for performing relatively unstructured queries which could not be performed by any other means, e.g. for material which is simply not indexed.
- **Ad-Hoc Search:** The Ad-hoc query facility allows authorized users to quickly create and perform sophisticated queries from anywhere within PowerCase. Queries or Searches can be performed *within* a particular incident, in which case the incident and any linked incidents will be searched; or *outside* of an incident, where the search will include all incidents across the PowerCase server that the user has permission to view.
- **Analysis:** this is concerned with the construction of queries to answer specific investigative questions. Such questions, paraphrased here, can be as broad as “Who knows who?”, and as specific as “Who knew the victim, is male, lives locally, has committed similar previous offences ...?”

These three facilities are described separately below.

Research

An important aspect of the whole indexing process is the avoidance of duplicate records. Research is therefore the process of taking all attributes values extracted into the index plan, and examining the database to see whether similar records already exist. The design objective for research in XANALYS PowerCase has been to make this process a single button press operation in the majority of cases whilst allowing the user to undertake further research if so desired.

Researching an index plan item can essentially be broken down into three distinct steps:

1. First the attribute values already present in the index plan are used to automatically populate a research query in an appropriate research form. Every attribute value extracted from the document is used in order to guarantee retrieval of all relevant existing records.
2. A database query is automatically generated by applying a Research Strategy to the attribute values supplied from the index plan.
3. Finally, a list of hits is prepared from the results of the query, ranked and presented to the user to select the best hit. The user has a choice of either to consolidating the new information into an existing record, or else to create a completely new record. XANALYS PowerCase also supports a compromise option, by which a new record can be created but linked by an "Identical?" link to existing records which could potentially be identical.

XANALYS PowerCase can utilize one or more pre-defined research strategies for each of the entity types. Each strategy defines the following:

- How attributes are to be grouped together, if at all. For example Forenames 1, 2 and 3 are grouped for cross searching. Height From and Height To are grouped in order to apply range matching. Many of the other attributes are handled independently.
- What match rules should be applied to each attribute group. For example for Surnames, Exact, Soundex, and Truncated Match Rules are applied. For the Forename group, Exact Synonym, and Truncated Match Rules are applied since forenames are not amenable to the Soundex algorithm, but are very commonly substituted by nick-names or shortened in some way.
- The positive and negative weighting to be assigned to each attribute group. These are used to accumulate an overall ranking score for each hit retrieved from the database. Separate Positive and Negative weighing values are used for successful and unsuccessful matches respectively. For example both the positive and negative weighting for the Surname of the person are relatively high since the surname, if queried, is a very significant element of the query. In contrast the positive weighting for the Sex attribute is quite low, as compared with that for surname, but the negative weighting is set quite high, since it is such a strong eliminating factor.
- How Reduction Factors should be applied between the different Match Rules employed on each attribute group. For example an Exact Match rule is usually given a reduction factor of 0%, whereas a Truncated Match Rule is more likely to attract a -15% or 20% reduction factor. The reduction factors, as their name suggests, are used to reduce the weighting for an attribute group.

- Last but not least the research strategy defines exactly what database tables and columns should be queried, and where appropriate what joins needs to be applied to construct the query.

All in all the research capability of XANALYS PowerCase is an extremely powerful facility, but one which goes somewhat unnoticed by many users, since it is quite intuitive. Without it however the user would be faced with the prospect of doing repeated queries of the database, each time changing options and where clauses in order to employ the alternative match rules.

Free text Search

XANALYS PowerCase also includes as free text search engine which can be used to conduct searches on textual material held in documents. This is regarded as important, complementary functionality to the Research capability described above and the analytical capability provided by the XANALYS Link Explorer product. For example the free text engine could be employed on, as yet, un-indexed documents, as a means of deciding which documents to index first. Alternatively it can be used to search for features which would not ordinarily be indexed.

Analysis – the Integration with XANALYS Link Explorer

XANALYS Link Explorer is another of Xanalys' Intelligence products. XANALYS PowerCase and XANALYS Link Explorer interwork with one another such that:

- XANALYS Link Explorer can be launched directly from XANALYS PowerCase
- XANALYS PowerCase can request that XANALYS Link Explorer generates a specific chart on a specific Query.
- XANALYS Link Explorer can request XANALYS PowerCase to open and Update Form on any object displayed in a chart.
- The user can quickly and easily switch between the two applications.

XANALYS Link Explorer is described in more detail elsewhere, but essential it is an extremely powerful data visualization and analysis tool, in which the user can construct arbitrarily complex queries through an intuitive, graphical user interface and visualize the results through a variety of chart type.

Document Workflow and User Authorization

Each user is authorized to play one or more roles such as a Typist, Indexer, Action Allocator, and Office Manager. Authorizing a User for a particular Role "enables" a set of Task Units, which in turn enables associated application functionality available through menus, dialogs and such like. In this way only authorized functionality is available to a user. Importantly, however, the user is authorized for the union of all Task Unit sets for all of their authorized roles. Thus the user is not required to switch roles continually whilst performing a mixed set of tasks.

XANALYS PowerCase incorporates a flexible work flow management system which controls the "flow" of documents through the various stages of processing. Tasks, such as "Type Statement S1" or "Index Statement S1", appear in Task Lists for each User Role. Each task corresponds to some operation that should be performed on some document or other object. Double clicking on a task in a Task List opens up the relevant document for editing, i.e. performing the operation.

When the user closes the document the user is prompted to set a new status for the document, which in turn is used to “Complete” tasks and possibly “Create” new ones.

Once a user has a form open, he or she is free to perform any operation on that document for which he/she is authorized. Importantly the user is not forced to open and close document repeatedly to perform separate operations. The only requirement placed on the user is the setting of the status value upon closing the document.

Case Linking and Triggering.

An important requirement in the Law Enforcement market is the ability to recognize and then manage a linked series of cases, the classic example being serial murder cases. XANALYS PowerCase supports what is termed as Case Linking, whereby two or more cases can be linked together as if they were one, but nonetheless still be maintained independently as separate cases. Case linking can be established between any pair of cases regardless of the servers on which the individual cases reside. Thus case linking is a form of distributed database management. One further feature of case linking is that the link between two cases need not be symmetric which thus permits different access privileges to be set up for different cases.

XANALYS PowerCase also supports a Triggering facility, whereby incremental comparisons are performed across all cases to determine potential coincidences or overlap between cases. The most obvious example of such a coincidence would be if the same person were the prime suspect in more than one case. However there are a whole spectrum of possibilities ranging from pure coincidences to subtle references to individuals, vehicles and locations.

New Features – PowerCase 4.4

Xanalys will release PowerCase 4.4 in Q1 2009. This release will include a range of features designed to reduce data entry overheads, expand search capabilities and speed up the process of data indexing:

- **Off-line Data Entry:** PowerCase’s document set will be mirrored by a set of off-line data entry forms, implemented using Microsoft InfoPath. These forms allow the officer to collect data in the field, and then upload the data directly into PowerCase, thereby removing the need to re-key the information.

These forms can be synched with the PowerCase database to ensure that the latest drop-down values are maintained.

Each PowerCase document follows a standard XML schema, meaning that agencies using different form applications (e.g. Adobe) can implement their own off-line document set.

- **Improvements to Ad-Hoc Search:** PowerCase’s Ad-Hoc search has been enhanced to include an improved user interface, the ability to include *optional* objects (database out-joins) and additional query test (NULL, NOTNULL).
- **Addition of E-Communications Object:** A new object type has been added that will capture E-mail, URL’s and other forms of electronic communications.
- **Index Plan Improvements:** The Index Plan has been enhanced to enable multi-object functions, such as delete, merge and X-referencing.

- Incident Administration Search: A new search facility within Incident Administration allows the user to quickly locate an incident and update its details, assignments, etc.

Xanals Limited

20-24 Church Street
Altrincham, Cheshire
WA14 4DW
United Kingdom

telephone +44 161 941 7792
fax +44 161 332 8285

www.xanals.com

