

# The Complete Guide to Eradicating PST Files



# The Complete Guide to Eradicating PST Files

Authored by Paul Cunningham, Dan Clark, and Thomas Madsen with a preface by Tony Redmond.

Published by QUADROtech

© Copyright 2015 by QUADROtech ([www.quadrotech-it.com](http://www.quadrotech-it.com))

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means without the written permission of the authors.

The example companies, organizations, products, domain names, email addresses, logos, people, places and event depicted herein are fictitious. No association with any real company, organization, people, domain name, email address, logo, person, place, or event is intended or should be inferred. The book expresses the views and opinions of the authors. The information presented in the book is provided without any express, statutory, or implied warranties. The authors cannot be held liable for any damages caused or alleged to be caused either directly or indirectly by this book.

**Please be respectful of the rights of the authors and do not make copies of this eBook available to others.**

First edition: 16 December 2015.

# Table of Contents

<b>Foreword</b> .....	<b>1</b>
<b>Preface</b> .....	<b>2</b>
<b>Chapter 1: The Problems with PSTs</b> .....	<b>4</b>
<i>Technical Issues</i> .....	4
<i>Security Issues</i> .....	6
<i>Compliance Issues</i> .....	7
<i>The Cost of PSTs</i> .....	8
<b>Chapter 2: Challenges in Eliminating PSTs</b> .....	<b>9</b>
<i>Stakeholder Support</i> .....	9
<i>Locating PSTs</i> .....	10
<i>Accessing PST Data</i> .....	10
<i>Establishing Ownership of PSTs</i> .....	11
<i>Re-ingesting PSTs</i> .....	12
<i>Finding a Balance</i> .....	13
<b>Chapter 3: Tools for PST elimination</b> .....	<b>14</b>
<i>Group Policy</i> .....	14
<i>Custom Scripting</i> .....	15
<i>Microsoft PST Capture Tool</i> .....	17
Preparing to Deploy PST Capture .....	17
Installing the PST Capture Console .....	18
Installing the PST Capture Agent .....	23
Running a Search with PST Capture .....	24
Importing PSTs with PST Capture .....	28
Issues with the PST Capture Tool .....	31
<i>Exchange Server PST Import Cmdlets</i> .....	32
Preparing for Exchange Mailbox Import Requests .....	33
Running Mailbox Import Requests with Exchange Server .....	34
Importing Multiple PSTs with Mailbox Import Requests .....	35
Removing Mailbox Import Requests .....	37
Issues with Mailbox Import Requests in Exchange Server .....	37

<i>Office 365 Import Service</i> .....	38
Migrating PST Data to Office 365 with the Network Transfer Method .....	39
Issues with the Office 365 Import Service .....	43
<i>SCANPST</i> .....	44
<i>Third Party Tools</i> .....	45
<b>The QUADROtech FlightDeck solution</b> .....	<b>46</b>
<i>Customer example</i> .....	46
<i>Approach</i> .....	47
Discover .....	47
Centralize .....	47
Process .....	48
Ingest .....	50
<i>FlightDeck Architecture</i> .....	52
<i>User Interaction</i> .....	54
<i>Owner Identification</i> .....	57
<i>Understanding the Return on Investment</i> .....	58
<b>Appendix: A Project Plan for Eliminating PSTs</b> .....	<b>60</b>

# Foreword

This book covers the topic of Personal Storage Tables (PSTs), how they are used in a Microsoft Exchange environment, some of the practical and operational difficulties that are encountered, and how the data contained in PSTs can be migrated to Microsoft Office 365 to become indexed and discoverable and so help companies attain compliance with the various legal and regulatory requirements that exist in different industries. These topics are addressed in the first three chapters.

The book is published by QUADROtech, a data transformation company based in Switzerland that has a long and proud record of helping customers move the data held in PSTs to Office 365. The last chapter contains a description of QUADROtech's FlightDeck technology and explains the process by which QUADROtech assists customers to locate and process PSTs.

The authors are:

Paul Cunningham, Exchange Server MVP and publisher of [ExchangeServerPro.com](http://ExchangeServerPro.com) and author of many books on Exchange and Office 365. Paul is available [on Twitter at @ExchServPro](#)

Dan Clark, Chief Strategy Officer of QUADROTech, and

Thomas Madsen, Senior Vice President of QUADROTech and product evangelist for FlightDeck.

Overall editing for the book and writing of the foreword is by Tony Redmond, author of fifteen books covering enterprise email since 1991 and Senior Contributing Editor for WindowsITPro.com, where he writes a [twice-weekly blog covering Office 365 and on-premises Exchange](#). You can follow Tony [on Twitter @12Knocksinna](#).

No editorial control over content written by Paul Cunningham or Tony Redmond was exerted by QUADROtech.

# Preface

***Tony Redmond***

A PST is a Microsoft proprietary file format designed to relieve the strain on expensive server storage by moving email items (messages, tasks, calendar items, and so on) and attachments to a local repository. Originally called a personal storage table to differentiate itself with the tables in the server-based Exchange online database, the PST is now often referred to as a personal storage file or personal folders. The PST format is similar to the OST (offline storage table or offline storage file) used by Outlook desktop clients to synchronize and cache slave copies of online folders for local use.

The need for PSTs was discovered soon after Microsoft launched the first version of Exchange server in 1996. Up to that time, the average size of an email was small (around 2 KB for a single screen of ASCII text). With the growing popularity of email as a method of corporate communication, message sizes rapidly grew, especially when multi-format attachments were factored into the equation, and the relatively small mailbox quotas assigned to users (often in the 20 MB to 50 MB range) were quickly exhausted, which then required the user to delete items to free space and allow new email to be delivered.

Outlook 97, the first release of the now-ubiquitous desktop client that connected to Exchange 5.0 and 5.5, popularized and accelerated the use of PSTs. At that time, it was common for email systems like Lotus cc: Mail, Microsoft Mail, and Lotus Notes to use local storage. PSTs provided a similar capability for users to move items from their online mailbox into their “personal storage” and so free space up in their mailbox without having to remove items permanently. In later releases, Outlook’s archive feature would detect and move old items automatically to the PST and so keep the online mailbox well under its quota. Outlook could also use the PSTs to receive new email, a feature used by many but one that I could never understand.

The PST file format is not secret and is [fully documented by Microsoft](#) to allow third-party developers to build tools that work against PSTs and led to the creation of some interesting products over the years. As discussed later in this book, many of the modern PST-centric products focus on the need to extract data from these files to move the information to a more accessible and secure location.

The original PST used an ANSI format and was restricted to 2 GB in size. However, users were unwise to allow their PSTs to grow to such a large size as corruption invariably resulted when the PST approached 2 GB. Problems with file structure and reliability were addressed by the introduction of the Unicode-format PST from Outlook 2003 onward. The new format was originally limited to 20 GB but [Outlook 2010 and later clients increase the limit](#) to 50 GB. The maximum size can be increased with a registry setting, which can lead to some very large PSTs. The largest I know of is some 62 GB and holds over 123,000 items. This monster is larger than the normal Office 365 mailbox (50 GB), so it cannot be ingested into a mailbox. However, it can be moved to an online archive, which is probably the more appropriate destination for PST data.

At one time, a compelling case could be argued to support the use of PSTs. Small mailbox quotas allocated to Exchange users and people often spent valuable time clearing out old messages by deleting them or moving them to a PST to allow new messages to be delivered. It's also true that network connections were incredibly primitive when viewed through the lens of today's always-on model. In those days, having data local meant that information was always at the user's disposal.

However, those reasons have disappeared into the fog of history. The average size of a corporate mailbox has expanded dramatically and is now often in the 5 GB to 10 GB range, with even higher amounts provided in cloud-based services like Office 365, where the default mailbox quota is currently 50 GB. In addition, archive mailboxes were introduced in Exchange 2010 as a way to provide online storage for items that needed to be retained for longer periods without necessarily being in the primary mailbox. One way of thinking about archive mailboxes is that they are a much smarter, more secure, online version of PSTs that the user doesn't have to manage. Network connectivity has improved dramatically around the world and depending on local access to data is no longer a necessity. In any case, experience proves that the vast majority of items moved into PSTs are never subsequently accessed, so they might as well be in an online archive anyway.

Modern storage is so cheap and plentiful that online capacity has moved to a point it is now "practically unlimited". Microsoft's introduction of [chained 50 GB chunks to form a single logical mailbox](#), a feature supported by both Office 365 and Exchange 2016 (soon), provides even the most prolific user with a way to move everything they store in PSTs today to an online repository. Once online, all of the advantages of high availability, security, and compliance are achieved.

Following high-profile cases such as the Sony hack in November 2014, when attackers penetrated and retrieved 179 PSTs containing some [73,000 messages from the mailbox of Sony Chairman Amy Pascal](#) along with other data, it's easy to understand why companies are reconsidering the use of PSTs. The need for organizations to comply with various legal and regulatory requirements has developed enormously in the last decade. PSTs were conceived at a time when the need for companies to be able to track, retain, and preserve email was not as great as it is now. If a company persists in allowing users to keep their PSTs, even to hold messages about long-gone subjects, it creates the potential of non-observance with a regulation (that might, for instance, require any communication relating to a topic to be kept for six years). Other potential problems include the inability to respond to law suits from disgruntled employees or to maintain control over intellectual property by being unable to prove that an idea or concept originated at a certain time.

Simply put, it's time to eradicate PSTs. Twenty years is too long to use a file format that is insecure and not very robust. It's time to move data online where it should be. Let's get on with the job.

**Stay tuned for the full book launching shortly...**