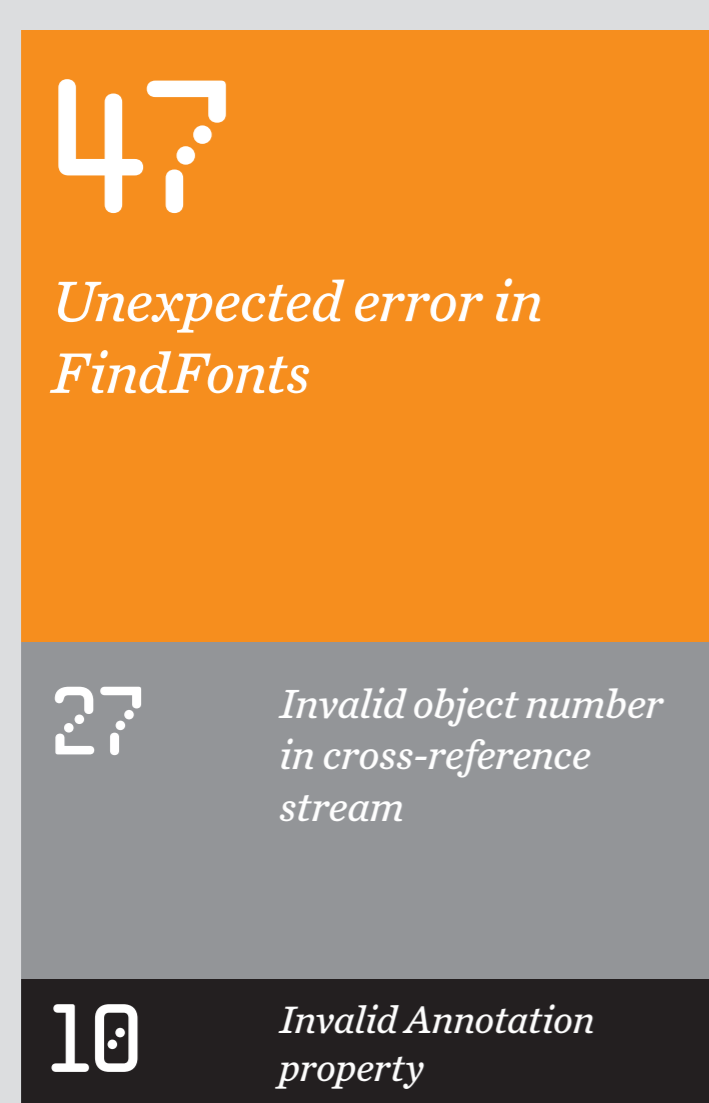# DIGITAL PRESERVATION

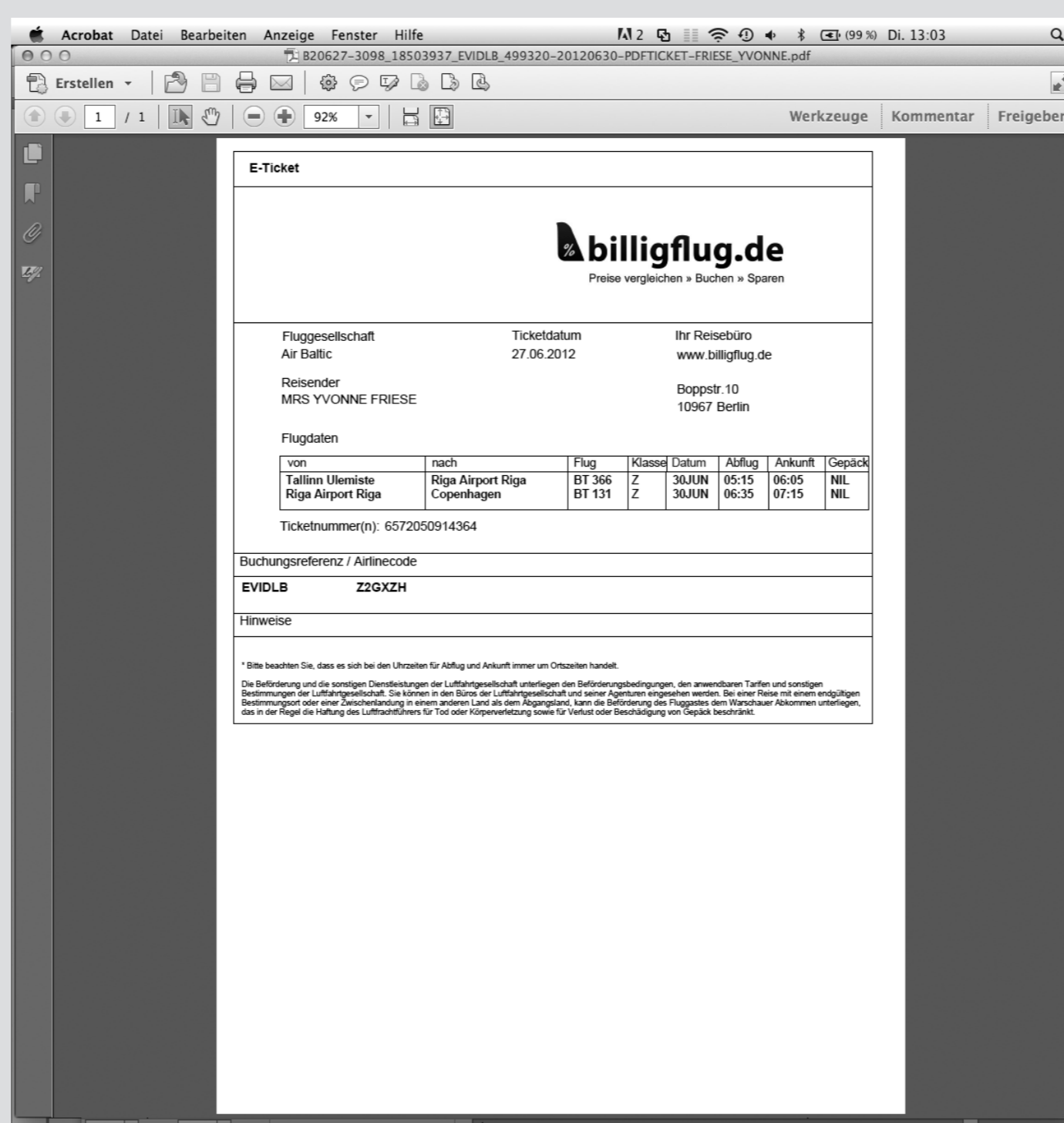*Hunger for Automation – The first migration actions in our Rosetta Digital Archive*

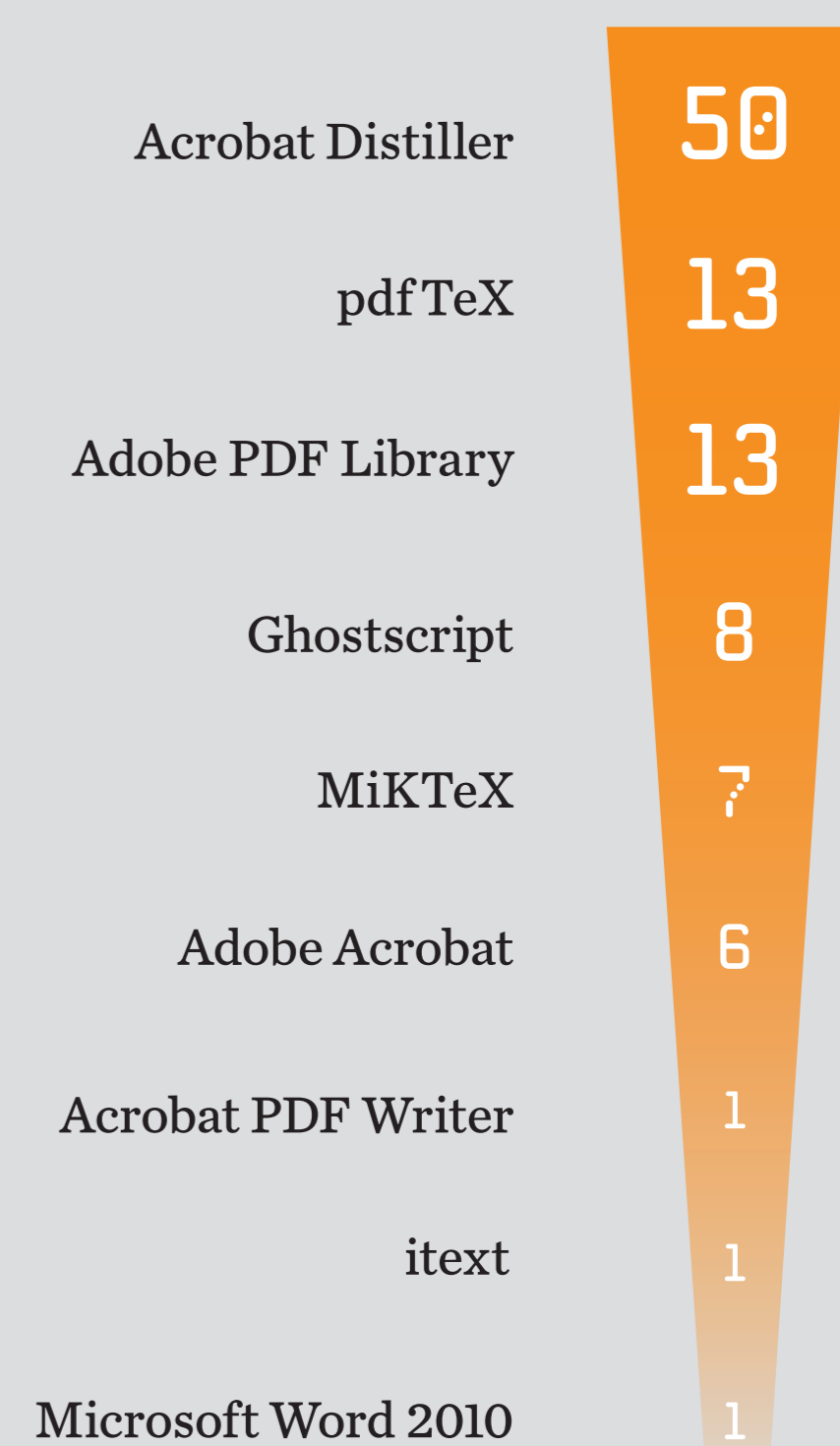## → CHALLENGE

Top 3 JHOVE Error messages *(in %)*

47 *Unexpected error in FindFonts*

27 *Invalid object number in cross-reference stream*

10 *Invalid Annotation property*

Example: Flight ticket in PDF Format – the information has to be reliable

Software used for PDF-Creation *(in %)*

Acrobat Distiller	50
pdf TeX	13
Adobe PDF Library	13
Ghostscript	8
MiKTeX	7
Adobe Acrobat	6
Acrobat PDF Writer	1
itext	1
Microsoft Word 2010	1

ZBW DSpace Open Access Repository *(in %)*

Ingest into the Archive 2012 : 47,000 PDF files

80 wellformed and valid PDF files

16 invalid and malformed

4 well-formed but not valid

Original Format

Tex Format **29%**

Microsoft Word **51%**

unknown **23%**
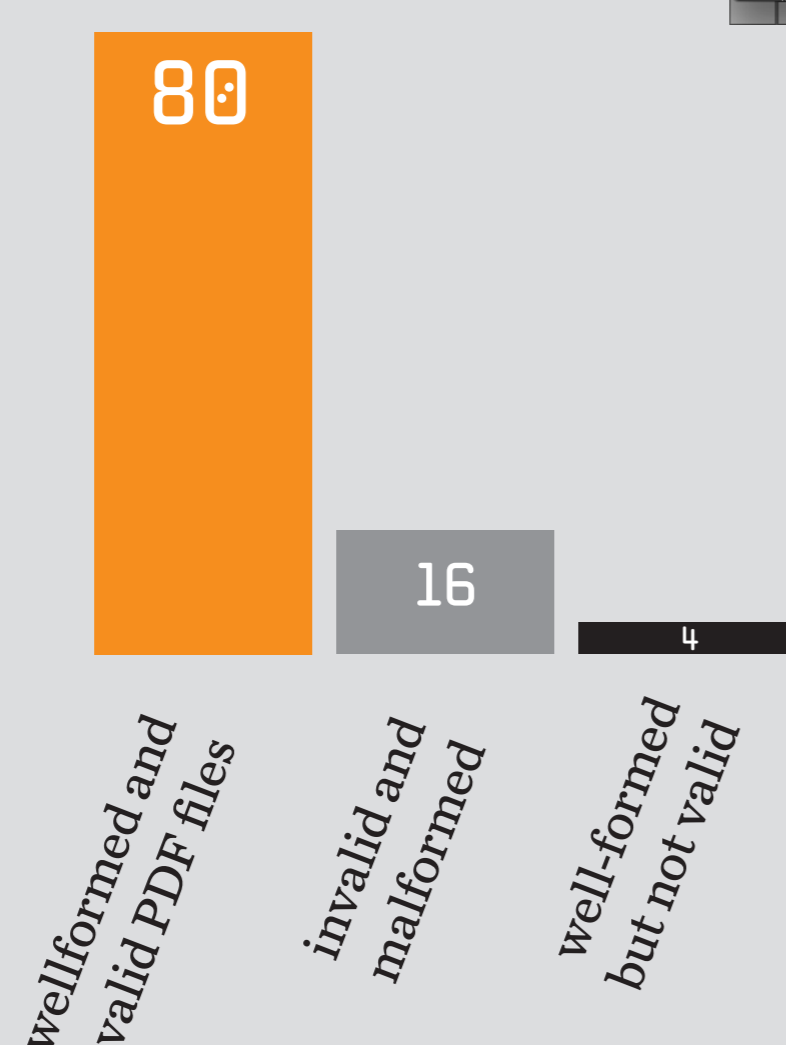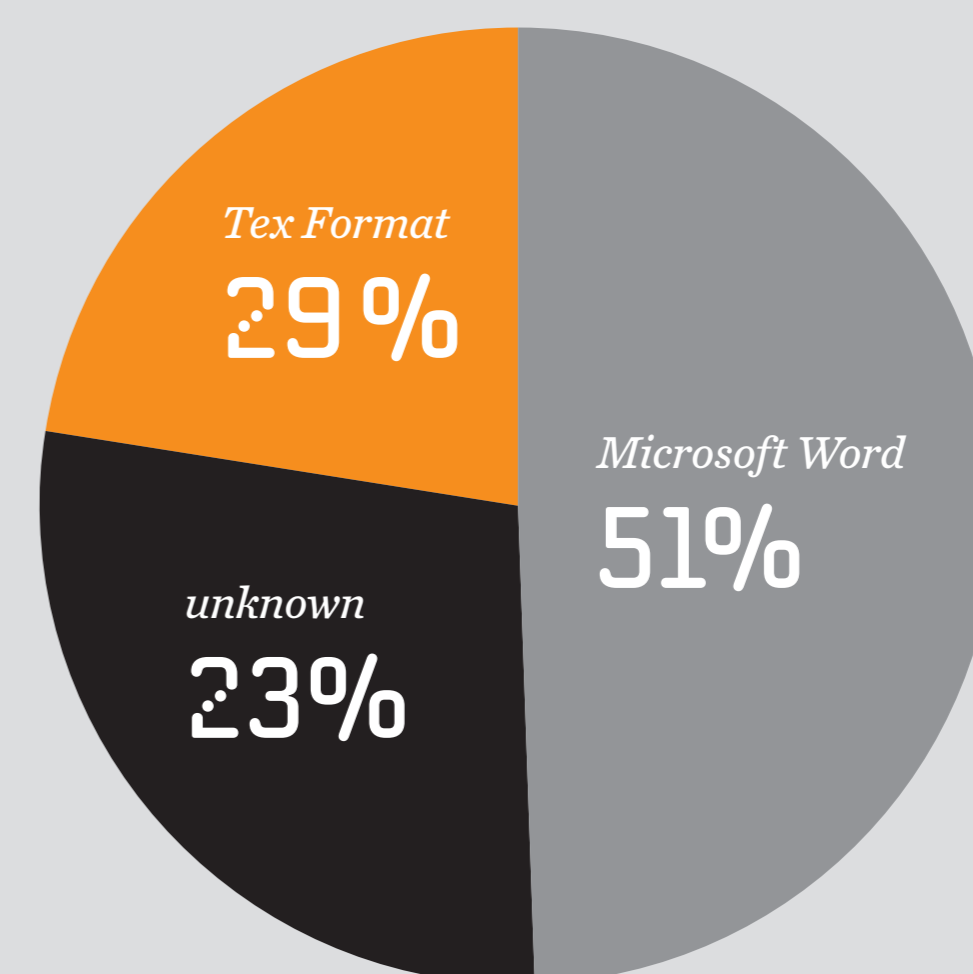
→ SOLUTION

Goal

– Migrate all invalid and malformed PDF-files to valid and well-formed PDF files version 1.5
– Keep the original PDF file and the migrated PDF within the same Intellectual Entity in the Digital Archive
– Have all steps documented in the Metadata

How to achieve it

– Via a self-built iText-based Java Tool
– Integrate the tool as a plugin for our Digital Preservation Software Rosetta
– Automate it for the PDF files that are already in the Archive
– Implement it into the workflow of the Preservation System to run it for all newly ingested PDF files as well