

The Dilemma between Technical Requirements and Simplicity -Future Proofing Rendering-

Derek Sergeant

12th March 2003

The problem

- Rendering a file
- Extracting information from a file
- Using a file
- Converting a file into a convenient format

The solution

- Remember what format the file is
- Maintain the tools that handle that format
- Create new tools that handle that format
- Maintain the associations between tools and formats

OAIS Representation Networks

- Use the same technology as the Cedars project to associate file formats with tools
- Allow for situations that are surprisingly different than those previously encountered

RepNet components

- Format specification
- Rendering tools
- Related formats
- Related tools
- Known issues
- Tools to aid in format identification

Two approaches

- Complete (or exhaustive) solution
- Pragmatic solution

Approach 1 - exhaustive RepNet

- Build a Representation Network for every conceivable instance of the file format
- Do this for every file format that the long-term archive will contain

Example - exhaustive approach

- Microsoft word file format

RepNet - word format object

Word format

Format specification

word for dos

v1 word for windows

v2

⋮

v10

Tools

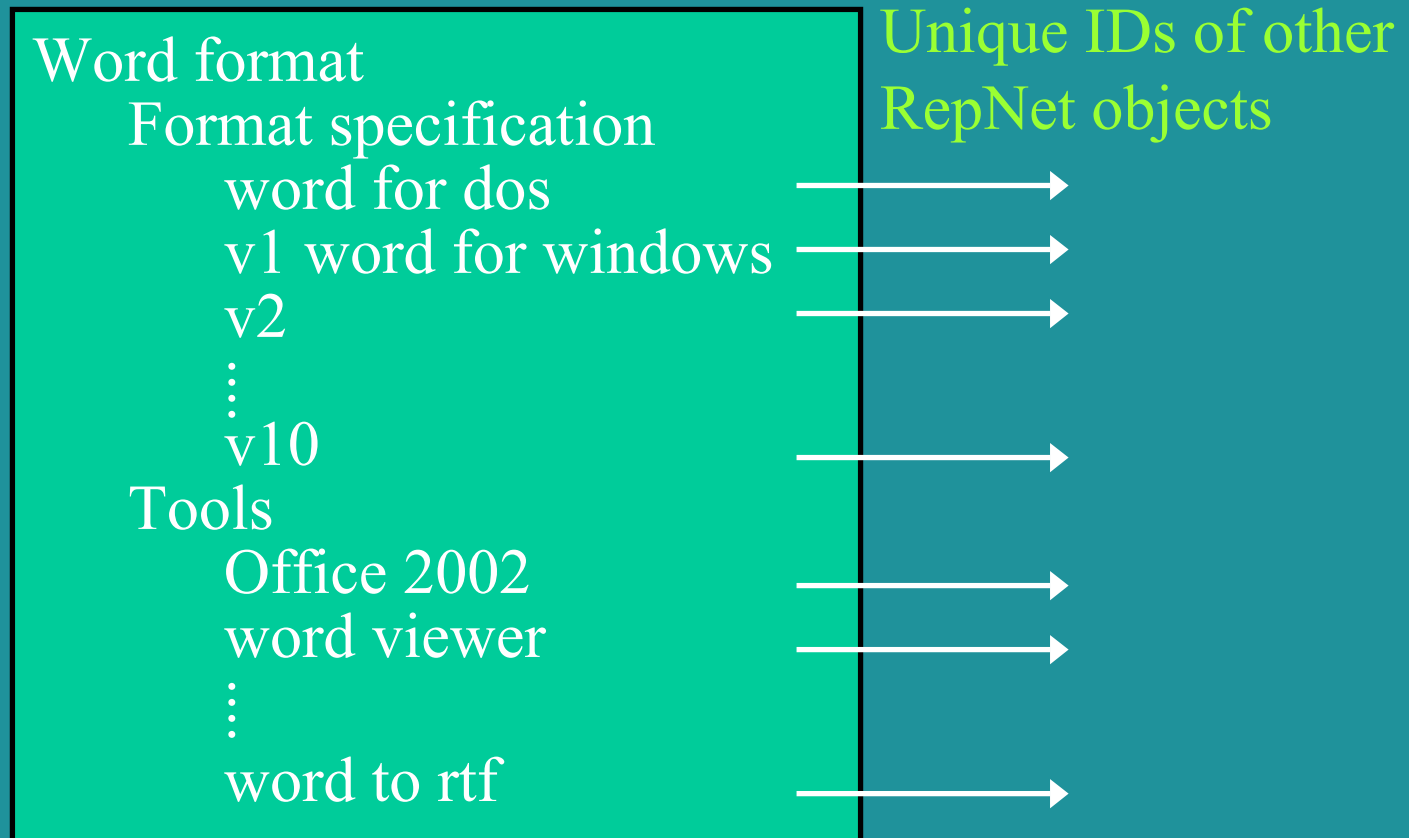
Office 2002

word viewer

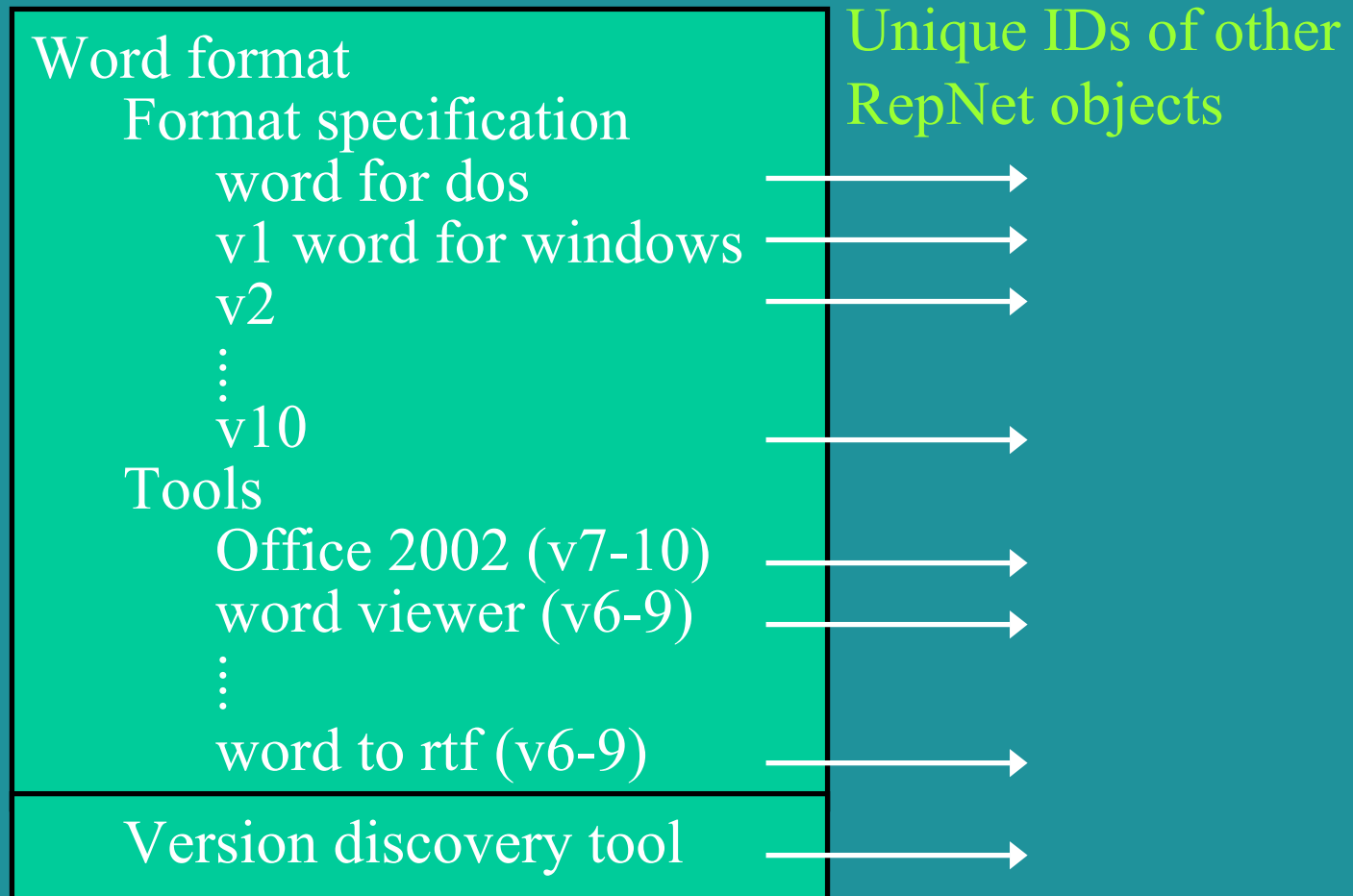
⋮

word to rtf

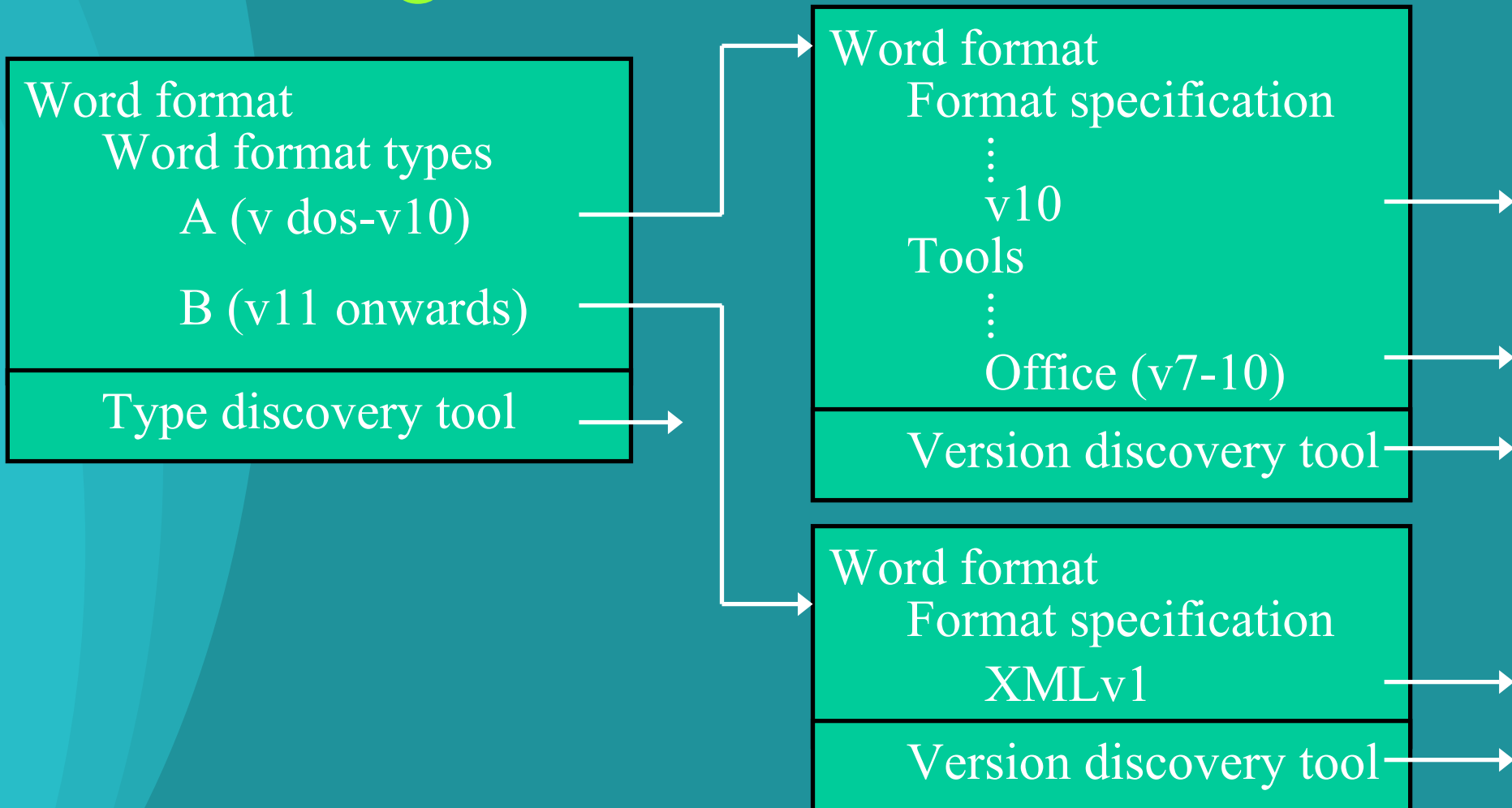
RepNet - word format object



Building up the word format obj



Building know-how of new versions



Approach 2 - pragmatic

- Microsoft word file format
- SVG graphics file format
- Mozart 3 music score file format

RepNet - word format obj

Word format

Informal description of the information we already have. If possible include references to external sources of information (i.e. DLF format registry)

RepNet - svg and mozart objs

SVG format

Informal description of the information we already have.

Mozart format

Informal description of the information we already have.

Pragmatic approach

- Keep track of which RepNets need attention
- Systematically improve these by priority

Update the word format object

Word format
Part of the microsoft office
suite. Has reached v10
(office XP). Further info @

Word format
Tools
word viewer →

Aide memoire
Part of the microsoft office
suite. Has reached v10
(office XP). Further info @

Update the SVG format obj

SVG format

Tools

Migration on Request

SVG viewer

Format specification

v2



Aide memoire

An XML based vector
graphics format. Good info

@ ... and @ ...

Update word format again

Word format

Tools

word viewer

word to rtf



Aide memoire

Part of the microsoft office
suite. Has reached v10
(office XP). Further info @

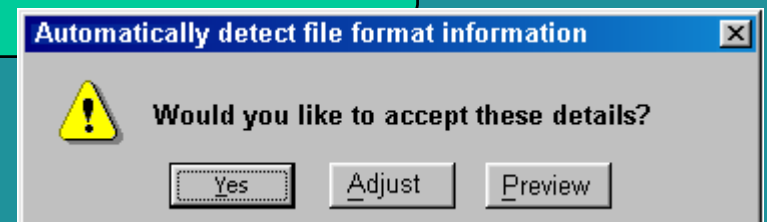
Simplicity for Joe Bloggs

- Complex information needs to be captured about a file format in order to render it
- Any extra information might prove useful in the future
- A RepNet should not discourage users from placing files into long-term preservation

A well designed interface

- Make the decision about “which file format?” easy
- Automate this and ask for confirmation

This is a “Word” document v8.03
There is Office 97 installed on your computer.
This will render successfully in WordViewer



Final thoughts

- Need sophisticated interplay between pragmatic (textual) and development of technology to convert informal description into complete RepNets
- Start with the most common or relevant formats

Final thoughts

- Media longevity is an issue, it is a mistake to let any deficiency in knowledge of a file format prevent or slow preserving the data
- The framework for recording file formats and RepNets needs to be flexible to allow situations that had not been thought of

Representation & Rendering Project

- Availability of file format information
- developing Migration on Request tools for text file formats
- designing and testing Representation Networks