

➔ **JPEG 2000 for long-term preservation in practice**

Problems, challenges and possible solutions

Johan van der Knijff

KB / National Library of the Netherlands

Contact: johan.vanderknijff@kb.nl

How this presentation came about

JPEG 2000 at the KB

2007/2008: Investigation of suitability JPEG 2000 for image masters (report Gillesse, Rog & Verheusen: *Alternative File Formats for Storing Master Images of Digitisation Projects*)

Early 2008: JPEG 2000 for mass-digitisation projects (preservation masters + access copies)

→ Actual digitisation outsourced to external companies

Ongoing digitisation projects

Project	# pages
Dutch parliamentary papers (1814-1995)	2.5 million
Dutch Prints Online (1781-1800)	2.1 million
Periodicals/ magazines (1840-1950)	1.5 million
Daily newspapers	8 million

Follow-up research (spring/summer 2010)

- Update of earlier research on adoption
- Software support
- Additional analyses on supplied image files
- Discussion of results with other institutions

Background: file formats

JP2 vs JPX

JP2

JPEG 2000 Part 1:

Basic still image format

JPX

JPEG 2000 Part 2:

**= JP2 + assorted
advanced stuff ...**

Colours: JP2 vs JPX

	JP2	JPX
Enumerated colour spaces	<ul style="list-style-type: none">• Greyscale• sYCC• sRGB	Many more, including CMYK
ICC Profiles	‘Restricted’ method: supports monochrome + 3-component matrix-based input profile	‘Any ICC’ method: supports any input profile

Resulting quality versus preservation conflict

- ICC profiles for working colour spaces (e.g. Adobe RGB, eciRGB) are typically '**display**' profiles (which are **not** 'input' profiles!!)
- So encoders that strictly adhere to the standard will **not** allow these colour spaces!
- Some encoders *do* allow them anyway → resulting files are not compliant with the standard!
- Applies to **both** 'Restricted' and 'Any ICC' methods!

JPX: preservation point of view

- Lots of complex features that we don't really want to deal with (multiple code streams, multiple compression methods, external references)
- Additional colour methods (incl. extended support of ICC profiles) that will be ignored by 'JP2-only' decoders
- Format is poorly supported by existing software

JPEG 2000 creation tools

“Big Three”:

- Aware JPEG 2000 command-line tool
- Luratech LuraWave command-line tool
- Kakadu command-line tools

Tiff → JP2 conversion Luratech

Issue:

Input TIFF contains ICC profile → Output JPX using
'Any ICC' method *

Risk:

- Loss of ICC profiles after future migration

*) Source: <http://jpeg2000wellcomelibrary.blogspot.com/2010/07/finding-jpeg-2000-conversion-tool.html>

Post-presentation update: apparently Luratech have added an option to change this behaviour in the latest (pre-release) versions of their software.

Tiff → JP2 conversion Kakadu

Issues:

- Capture resolution written to ‘display resolution’ header field in JP2
- ICC profiles not preserved after conversion

Risks:

- Loss of resolution info after future migration
- Loss of ICC profiles (*now!*)

Adobe Photoshop (Adobe plugin 2007, CS4)

Issues:

- Output always JPX ('JP2-compatible' optional)
- ... but 'brand' header field is always JP2!!

Risks:

- Unpredictable software behaviour (files are not what they purport to be!)
- Loss of ICC profiles after future migration

Tiff → JP2 conversion ImageMagick

Issues*:

- ICC profiles not preserved after conversion
- No support of JP2 resolution box
- Performance

Risks:

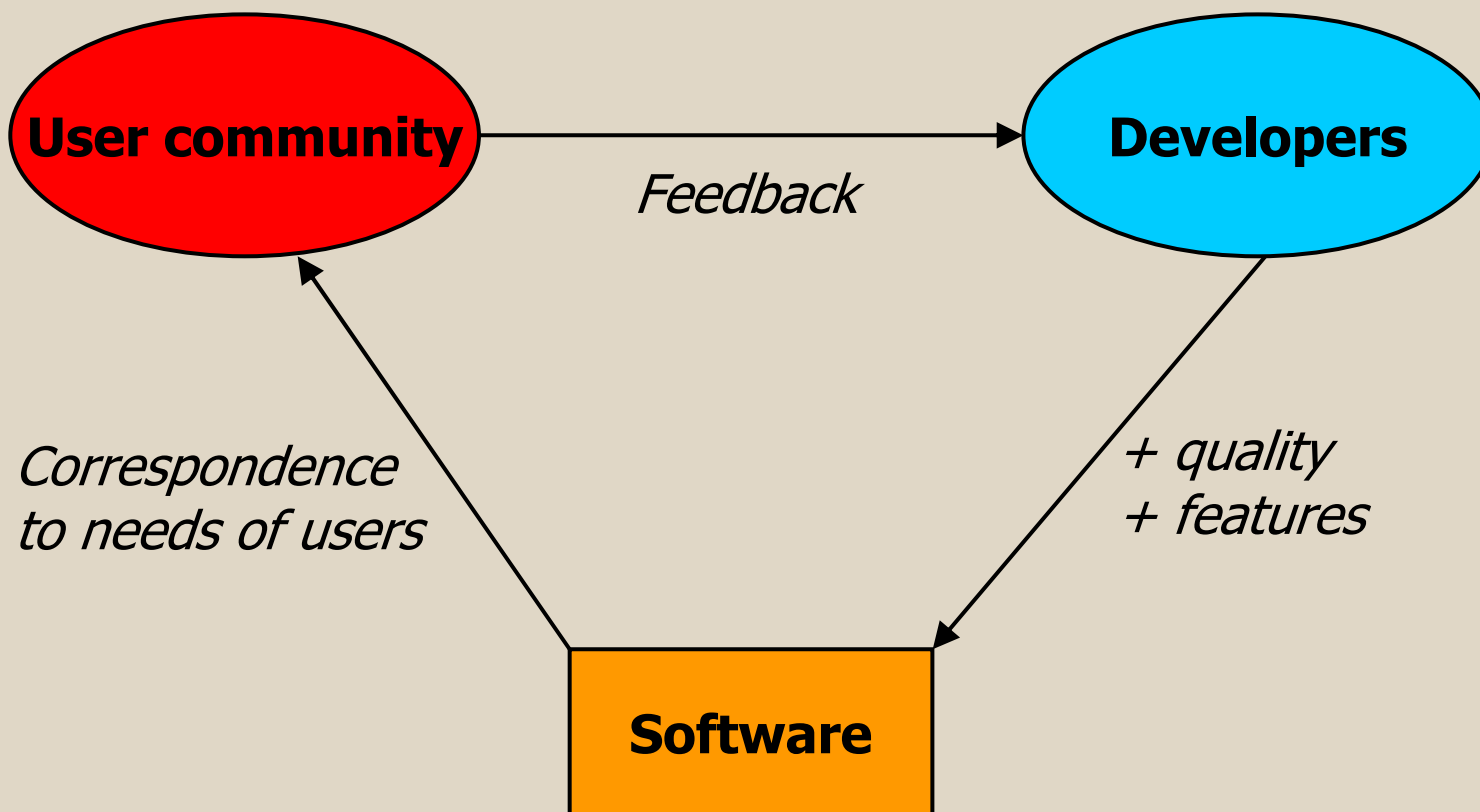
- Loss (*now!*) of ICC profiles
- Loss (*now!*) of capture resolution info

*) All caused by underlying *JasPer* library, not IM itself!

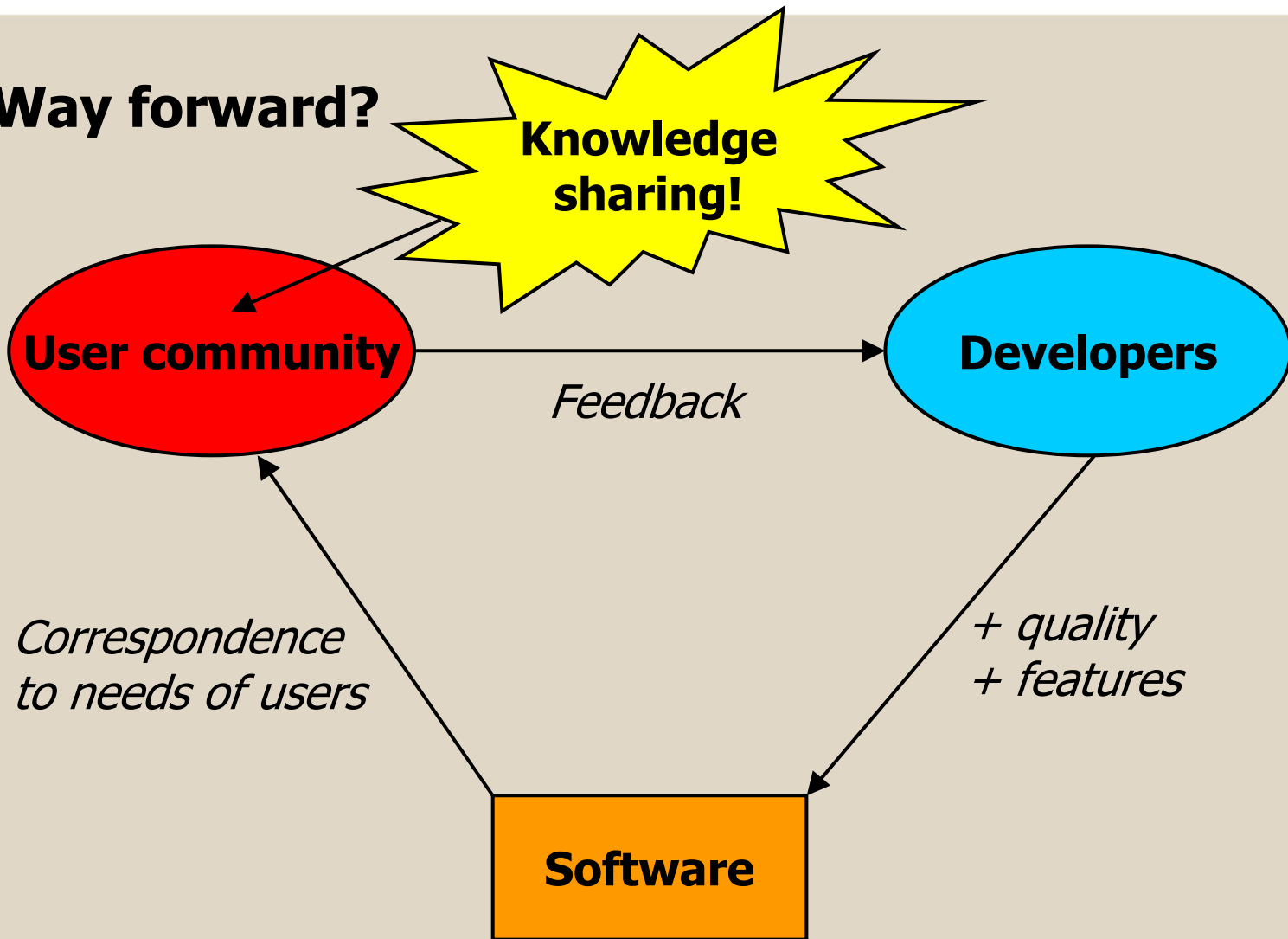
Creation tools wrap-up

- Number of available tools is limited
- Issues and mis-interpretations of filespec are still common
- Even minor issues can have a large impact for long-term preservation (repositories may contain *millions* of images; images may have been created by various tools)!

Way forward?



Way forward?



Rob Buckley: amendment JP2 file spec

- Meet the needs of the Cultural Heritage Community
 - Align JP2 with ICC practice for RGB color spaces
 - Provide migration path for existing TIFF content
 - Upgrade to latest ICC spec: ISO 15076-1:2010
- Proposal presented last summer at JPEG meeting in Brussels: support of full ICC specification
- May be sufficient to remove input profile class restriction, ICC supports this

Software libraries

Open source

JasPer

- current stable version early 2007
- incomplete JP2 support, poor performance

JJ2000

- not updated since 2002

OpenJPEG

- current stable version late 2007

Implications for developers

Support of JP2 in existing software obstructed by lack of robust, high-performance libraries



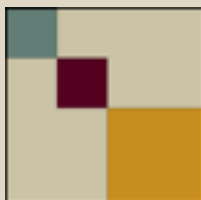
ImageMagick



IIPImage



GIMP



Djatoka

Implications for users

- Keeps choice of reliable encoding/decoding options limited
- Dependence on limited number of (sometimes expensive) commercially available tools
- What if one of the 'Big Three' goes out of business, or decides that JPEG 2000 product line is no longer profitable?

Way forward?

We would need a JP2 library that is:

- Feature-rich: encode, decode, reformat
- Compliant with JP2 file specification
- High-performance
- Open source
- Along lines of LibTIFF (TIFF library)

→ Proposal Rob Buckley: needs funding, project partners

JP2 as a preservation format

“What’s the problem? So many other archival institutions are already using JP2/JPEG2000!”

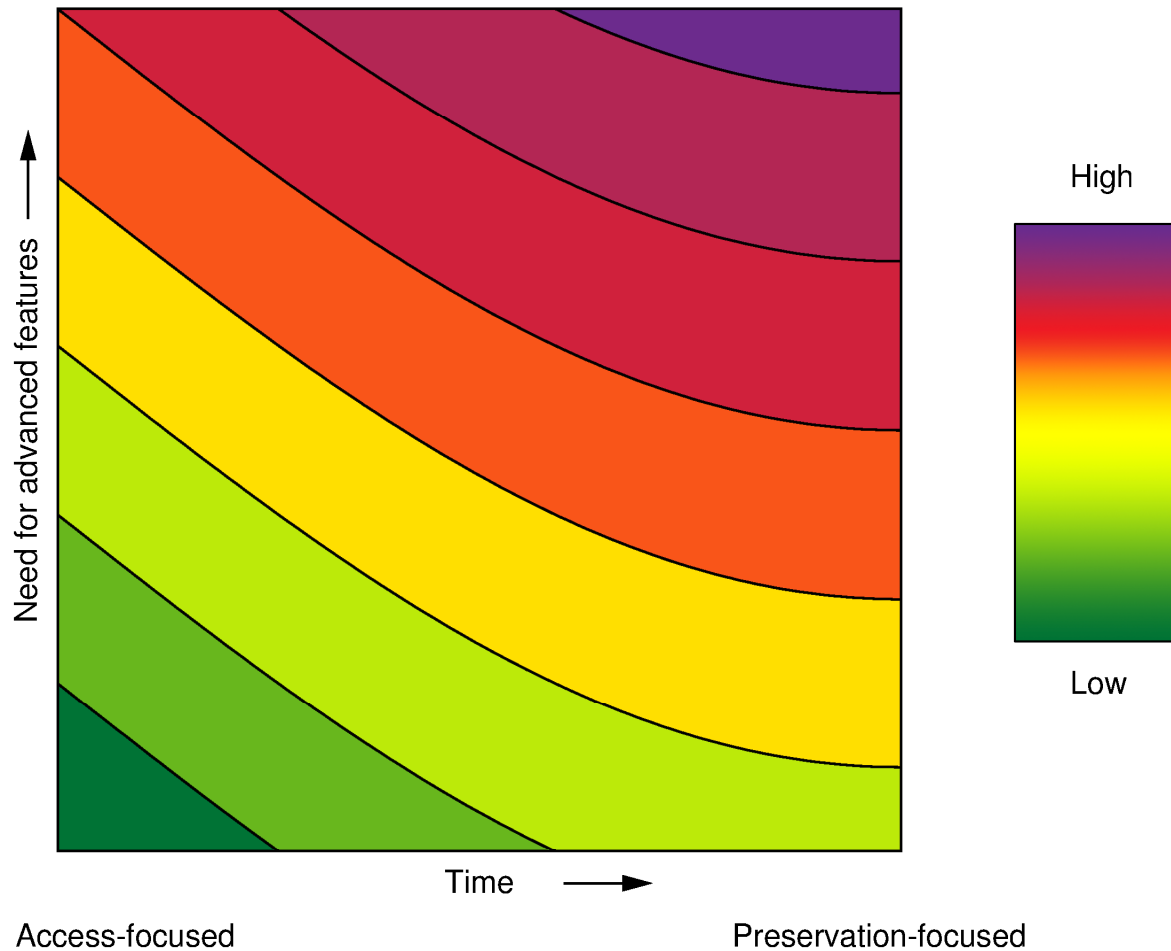
?

Doesn't say much without taking into account:

- Purpose: long-term preservation <--> short-term access?
- Quality requirements?
- Awareness of limitations/quirks of specific software?
- Awareness of resulting risks?

Implementation difficulty (\propto preservation risk)

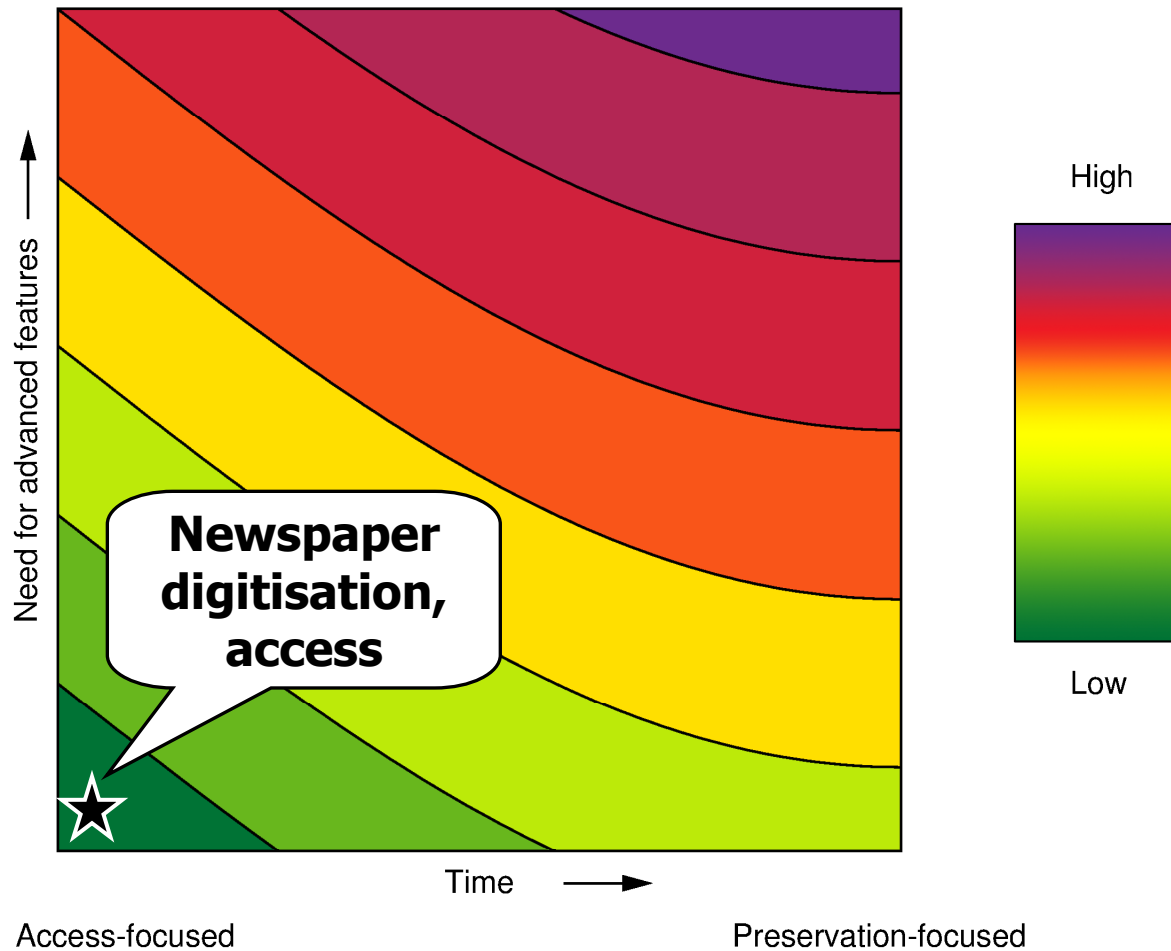
ICC-profiles,
embedded
metadata etc.
important



Only image
codestream
really matters

Implementation difficulty (\propto preservation risk)

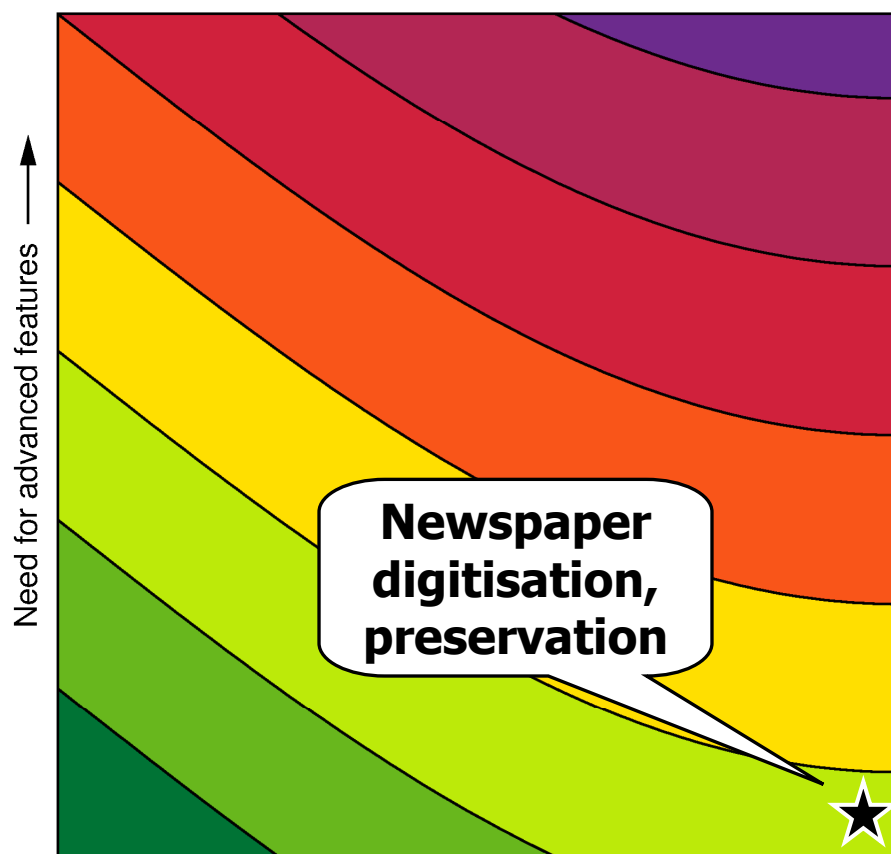
ICC-profiles,
embedded
metadata etc.
important



Only image
codestream
really matters

Implementation difficulty (\propto preservation risk)

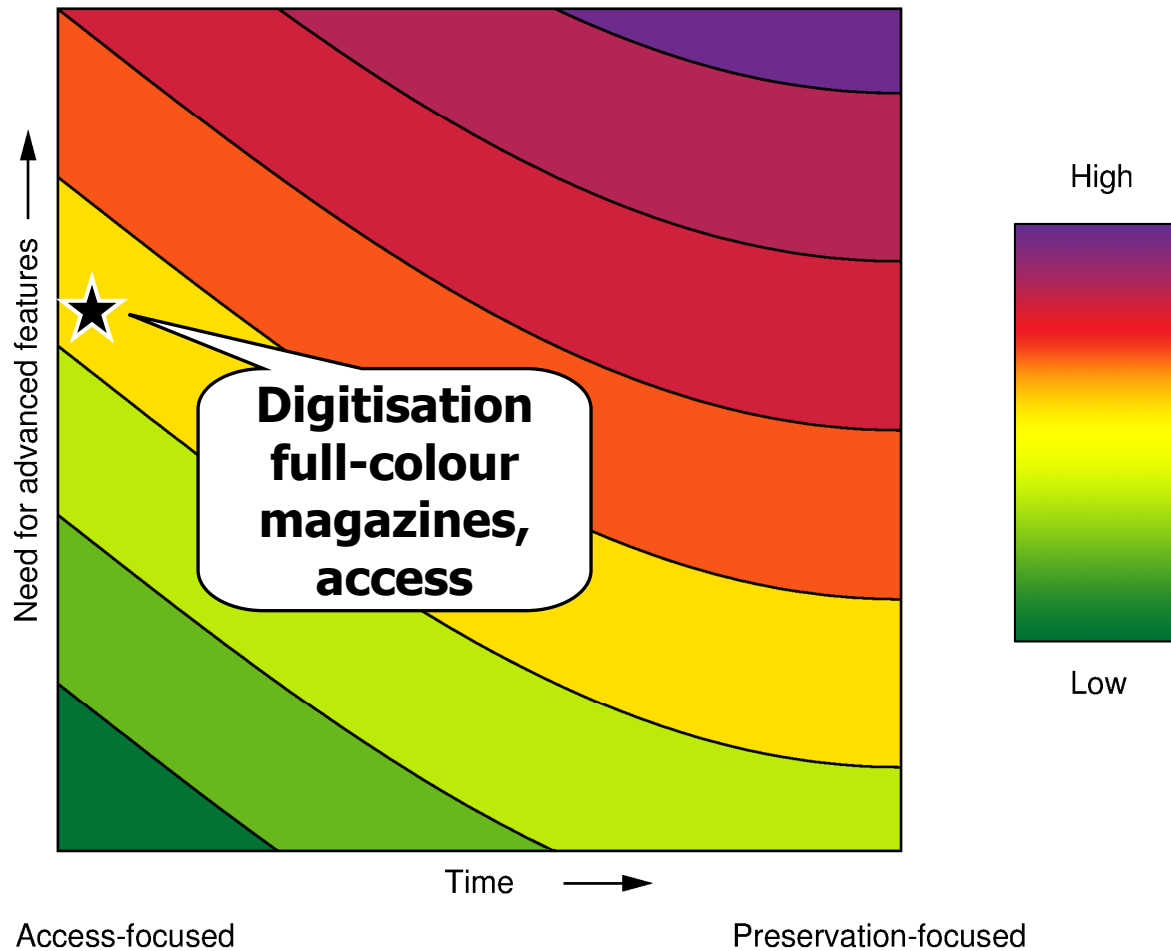
ICC-profiles,
embedded
metadata etc.
important



Only image
codestream
really matters

Implementation difficulty (\propto preservation risk)

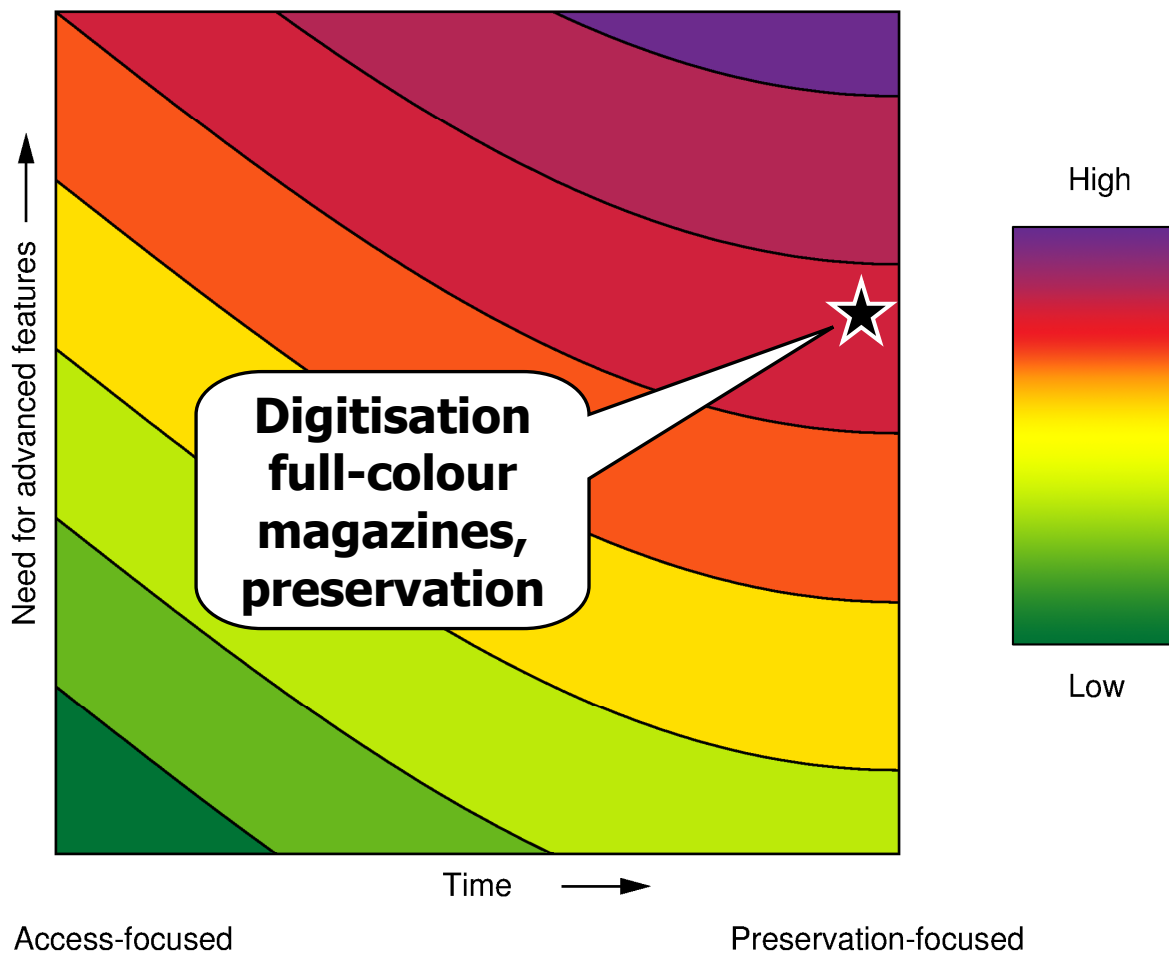
ICC-profiles,
embedded
metadata etc.
important



Only image
codestream
really matters

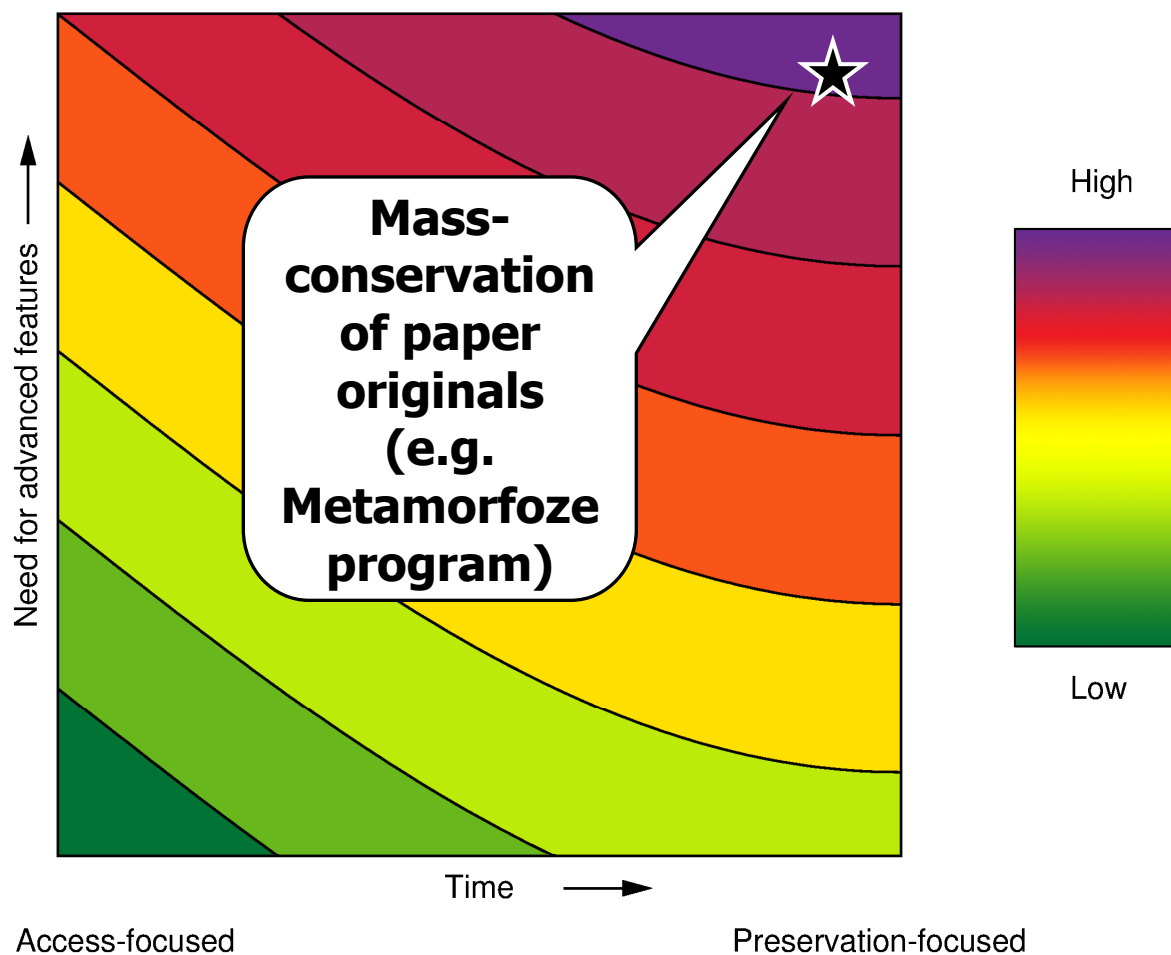
Implementation difficulty (\propto preservation risk)

ICC-profiles,
embedded
metadata etc.
important



Implementation difficulty (\propto preservation risk)

ICC-profiles,
embedded
metadata etc.
important



Wrap-up

Software

- Issues in existing JPEG 2000 creation tools introduce (hidden) preservation risks!
- Sharing of knowledge on these issues is important
- Feedback to software manufacturers is important
- Limited availability of encoding/decoding options introduces risks
- Development of a stable, feature-rich high performance open JP2 software library could be a major step forward → **needs action from user community!!**

Suitability of JP2 for preservation

- Suitability of JP2 (or any other format) depends on specific purpose/aim (e.g. access, preservation, replacement)
- Quality requirements are important (e.g. colour support)
- Current filespecs do not allow proper colour management → amendment/correction of ICC support required to meet needs of preservation community
- General statements such as “*Format X is suitable / not suitable for preservation*” don’t say much and are not very helpful!