



**PDF Days Europe 2022 | Berlin**

# Combining PDF 1.7 and 2.0

Driving Accessibility, Reflow and Reuse

Matthew Hardy | Director of Engineering | Adobe

“This project’s goal is to solve a fundamental problem that confronts today’s companies. The problem is concerned with our ability to communicate visual material between different computer applications and systems.”

*The Camelot Project*  
*J. Warnock*











of an

Adapt to the

Registrar's Office

Rank or Profession

Residence at the time of Marriage

ENTRY OF MARRIAGE

Marriage Ordinance

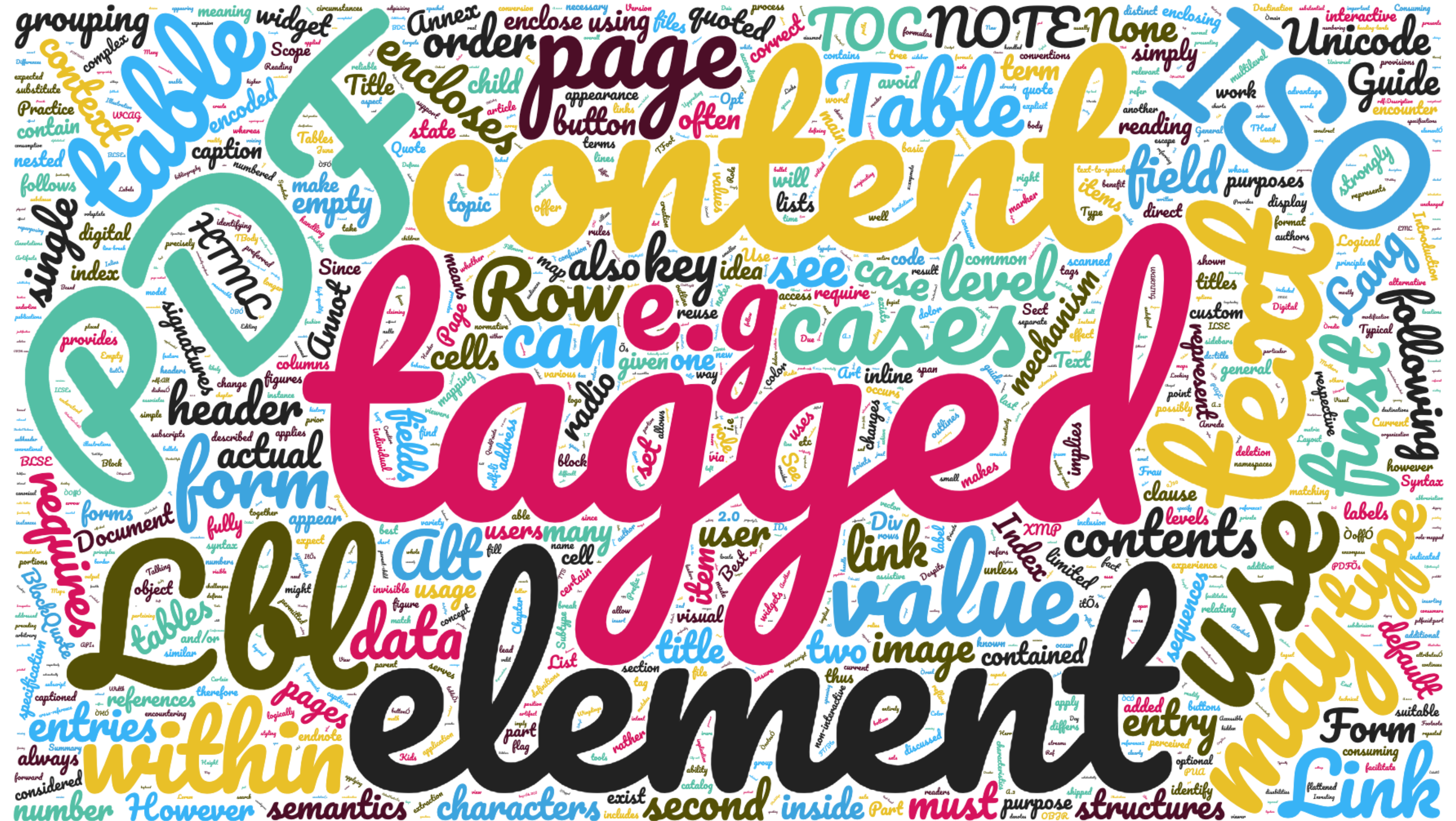
Entry No. 21

an











# Tagged PDF



## Tagged PDF

- 21 years in PDF (PDF 1.4)
- Broad use cases

## Solves for

- Accessibility
- Content Reuse
- Reflow
- Extraction





# ISO 32000-2: Tagging Update



## PDF 1.7

- Complexity to tagging (imprecise rules)
- Inability to interchange complex tagging — no namespaces

## PDF 2.0

- New and improved (additional) tag set
- Reduce complexity and provide increase interoperability
- Namespaces to enable complex interchange
- Both PDF 1.7 and PDF 2.0 types available
- How and when to use both — challenge



# PDF 1.7 and 2.0 namespaces



## PDF 1.7 and 2.0

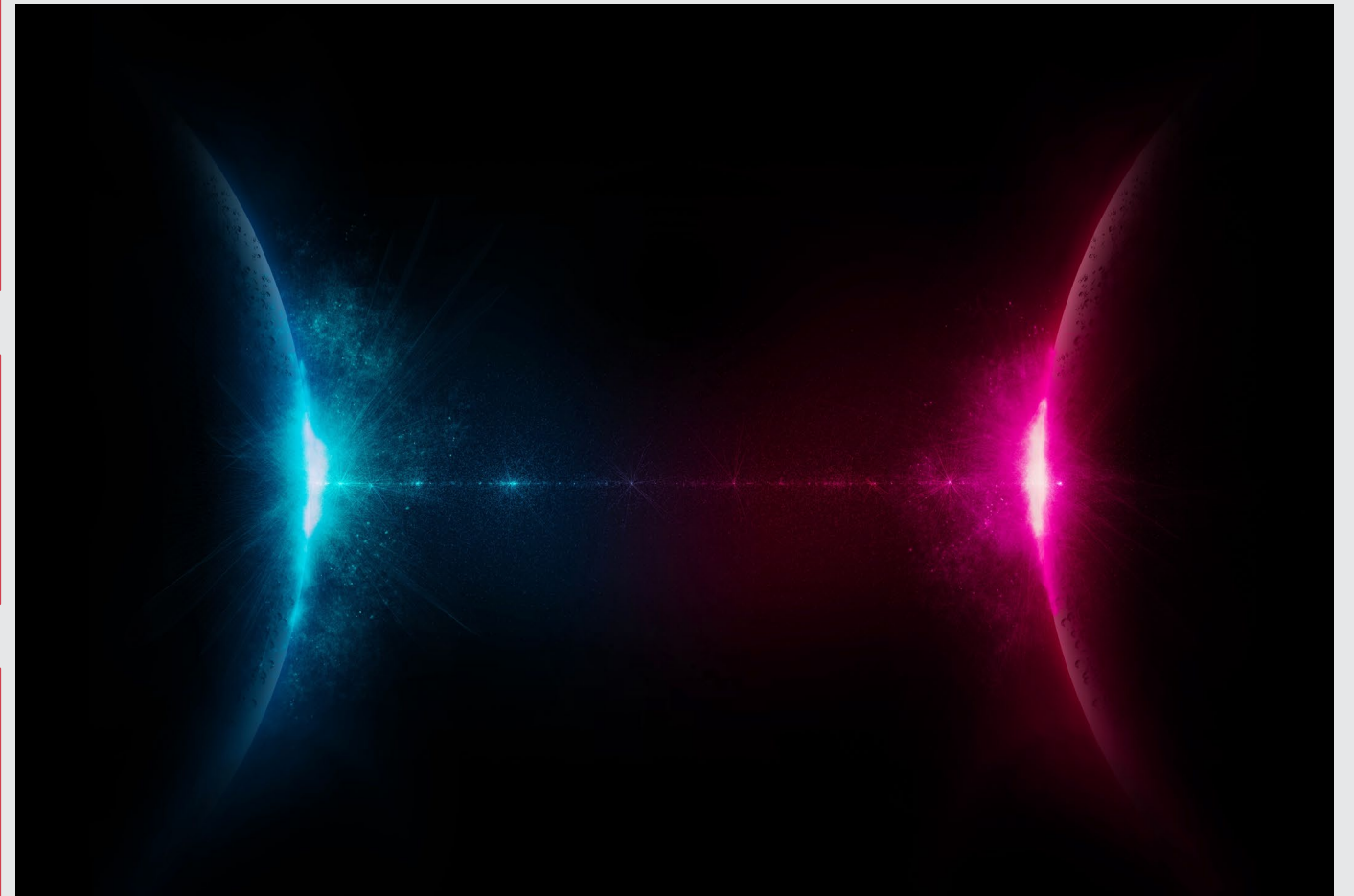
- Namespaces for both defined in PDF 2.0
- Co-existing
- Both are usable
- Interaction between two undefined

## PDF 2.0 Namespace

- Cleaner definition of structure types
- Guidance on usage
- Annex L provides “schema” for PDF 2.0 namespace

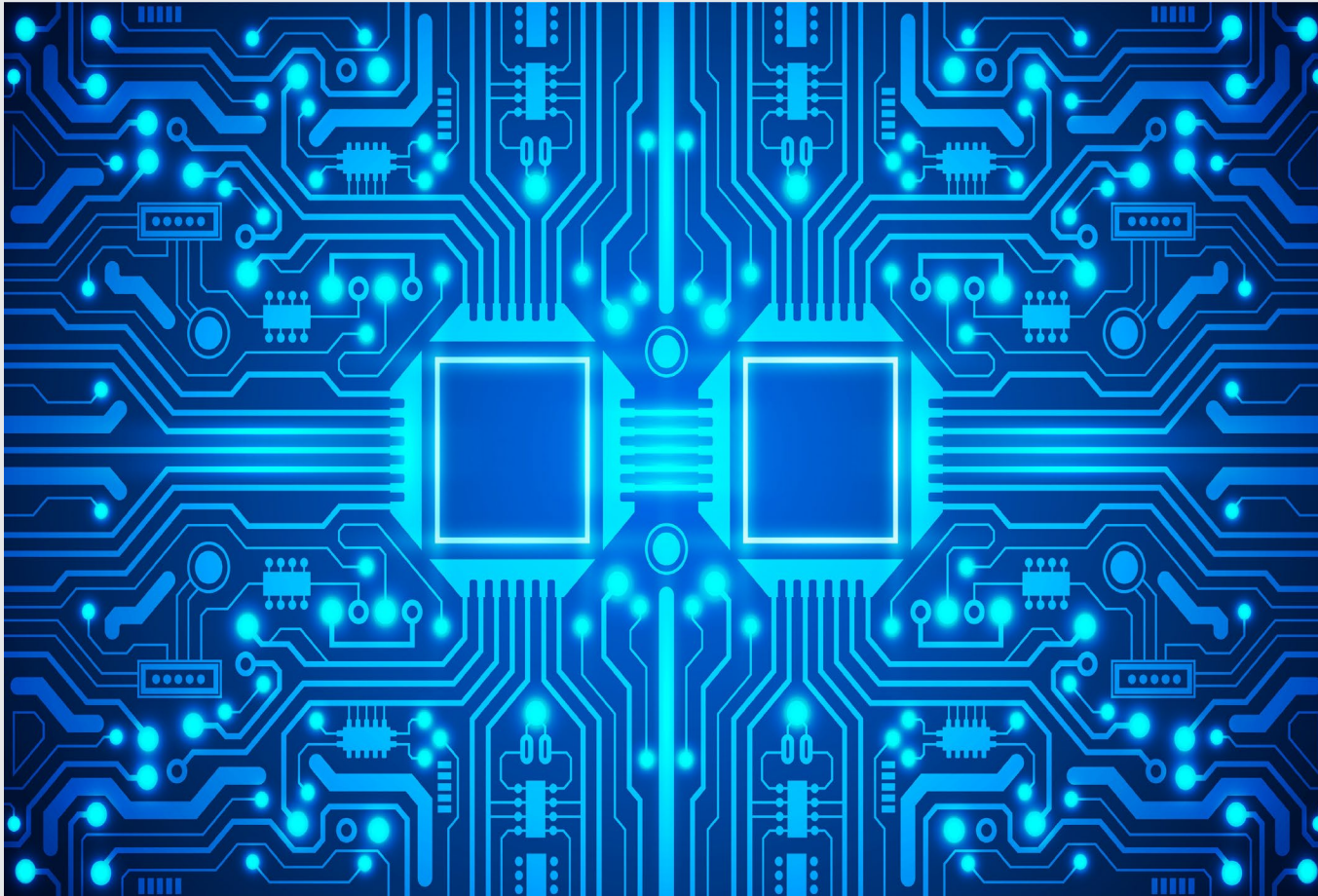
## PDF 1.7 Namespace

- Loosely defined (references 32000-1)
- No schema or clarified definitions
- Unclear how PDF 2.0 attributes relate to PDF 1.7





# 32005: Unifying PDF 2.0 and PDF 1.7



## PDF 1.7 namespace

- Cleaner definition of unique PDF 1.7 elements
- Examples of usage and intent
- Ensures namespaces are used for clarity

## PDF 2.0 Attributes

- Superset of PDF 1.7
- Applicable to both namespace equally
- All elements can and must use PDF 2.0 attributes

## Interchange

- ISO 32005 allows a “pure” PDF 1.7 document to be updated
- Any PDF 2.0 tagging features guarantees use of namespaces
- Marker to identify PDF 2.0 documents conforming to 32005



# ISO 32005: Schema (Mega-table)



## ISO 32000-2 (PDF 2.0)

- Schema for PDF 2.0 tags
- Intentionally ambiguous for PDF 1.7 namespace

## ISO 32005

- Complete schema across the PDF 1.7 and PDF 2.0 namespaces
- Use PDF 2.0 definitions when element types are common
- Enable unique PDF 1.7 elements in the PDF 2.0 namespace
- Exclusion for pure PDF 1.7 tags





# 32005: Outcome



## Expanded Semantics

- Common misunderstanding that PDF 1.7 elements were not permissible in PDF 2.0
- Clear guidance on how to include the PDF 1.7 namespace in PDF 2.0 document
- No loss of richness of semantics

## Compatibility

- ISO 32005 is completely compliant with ISO 32000-2
- It is NOT an extension
- Clarifies ambiguity in PDF 2.0 documents
- Available for future standards built on top of 32000-2



## Gold standard for Accessibility

- High-quality tagging
- Part 1 ignored all but visual and basic semantics
- Part 2 leverages 32005 and enables broader use cases

## Use cases

- Screen readers
- Magnification
- Adaptive contrast
- All through tagged PDF and 32005



# Focus on Accessibility



## Tagged PDF ecosystem

- Highly focused on accessibility

## Tunnel vision

- Regulatory requirements around accessibility
- Meeting minimum requirements for accessibility
- Other uses primarily ignored



# Tagging Gap



## Tagged PDF

- Presence of tags insufficient to guarantee quality

## Minimal use of Tagged PDF

- Inconsistent use of tags
- No attributes
- Limited information
- Primarily aimed at text extraction



# Well-Tagged PDF (WTPDF)



## Features

- Mandate optional tagged PDF features and their use
- Optimize for interchange
- Built on ISO 32005

## Focus

- Content Reuse
- Reflow
- Extraction

## Clear identification

- WTPDF marker in the PDF
- Clearly identify high-quality tagged PDF



# 32005, WTPDF & PDF/UA



## 32005, PDF/UA-2 and WTPDF

- Aligned by design
- 32005 provides schemas
- UA and WTPDF extend

## WTPDF

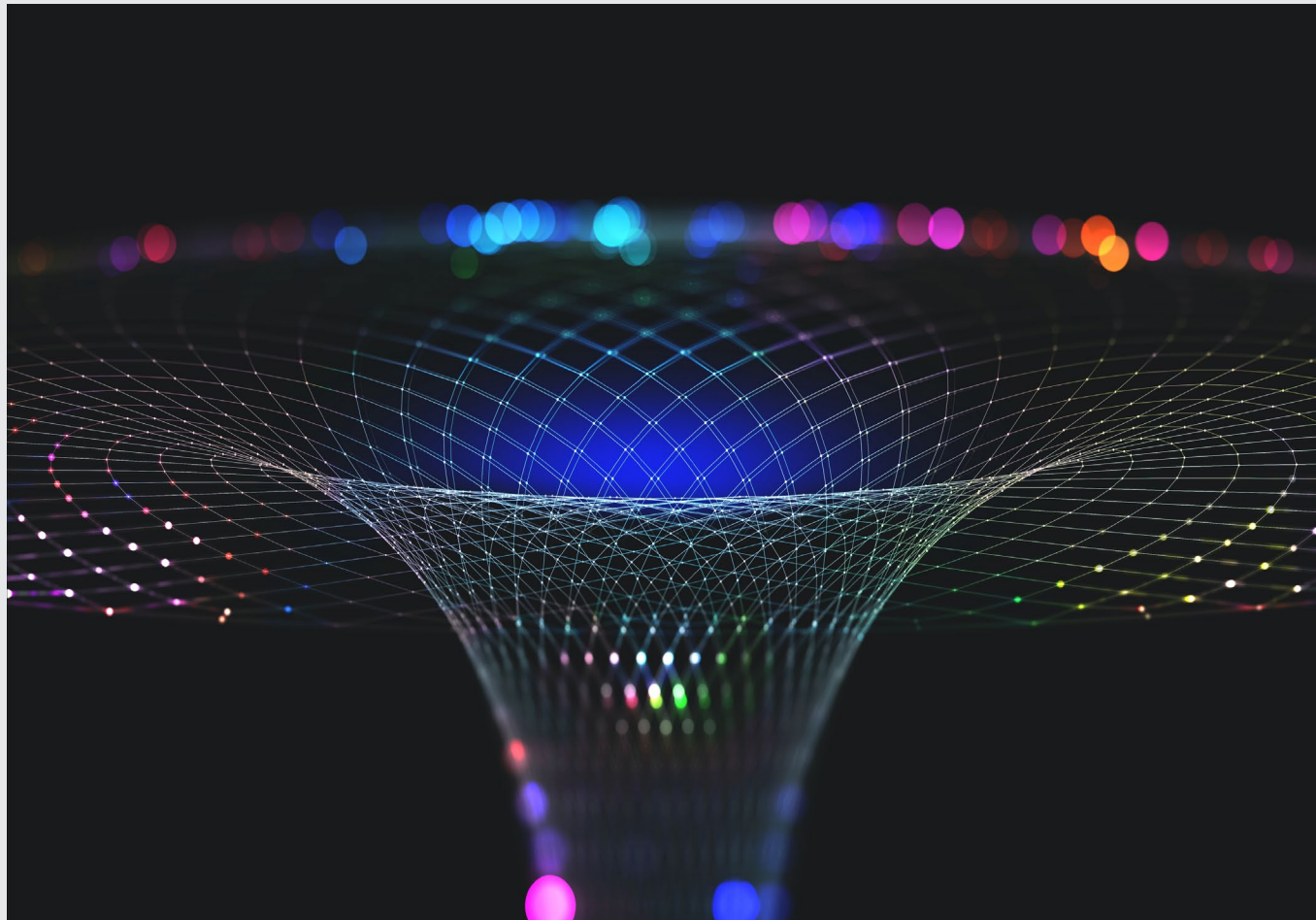
- Output from producers and converters
- Enables future for PDF

## PDF/UA

- Brings Accessibility
- Best in class with WTPDF marker



# HTML Derivation



## Derivation TWG

- Defines industry standard for PDF to HTML conversion
- Powered by Tagged PDF

## 32005 & WTPDF

- Derivation is a strong use case for these standards
- 32005 validated for compatibility (and adapted)
- WTPDF intended to ensure information for derivation



# Opportunities



Modern device support



Experiencing documents



Unlocking data in PDF



Many opportunities