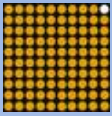


# BeyondRecognition

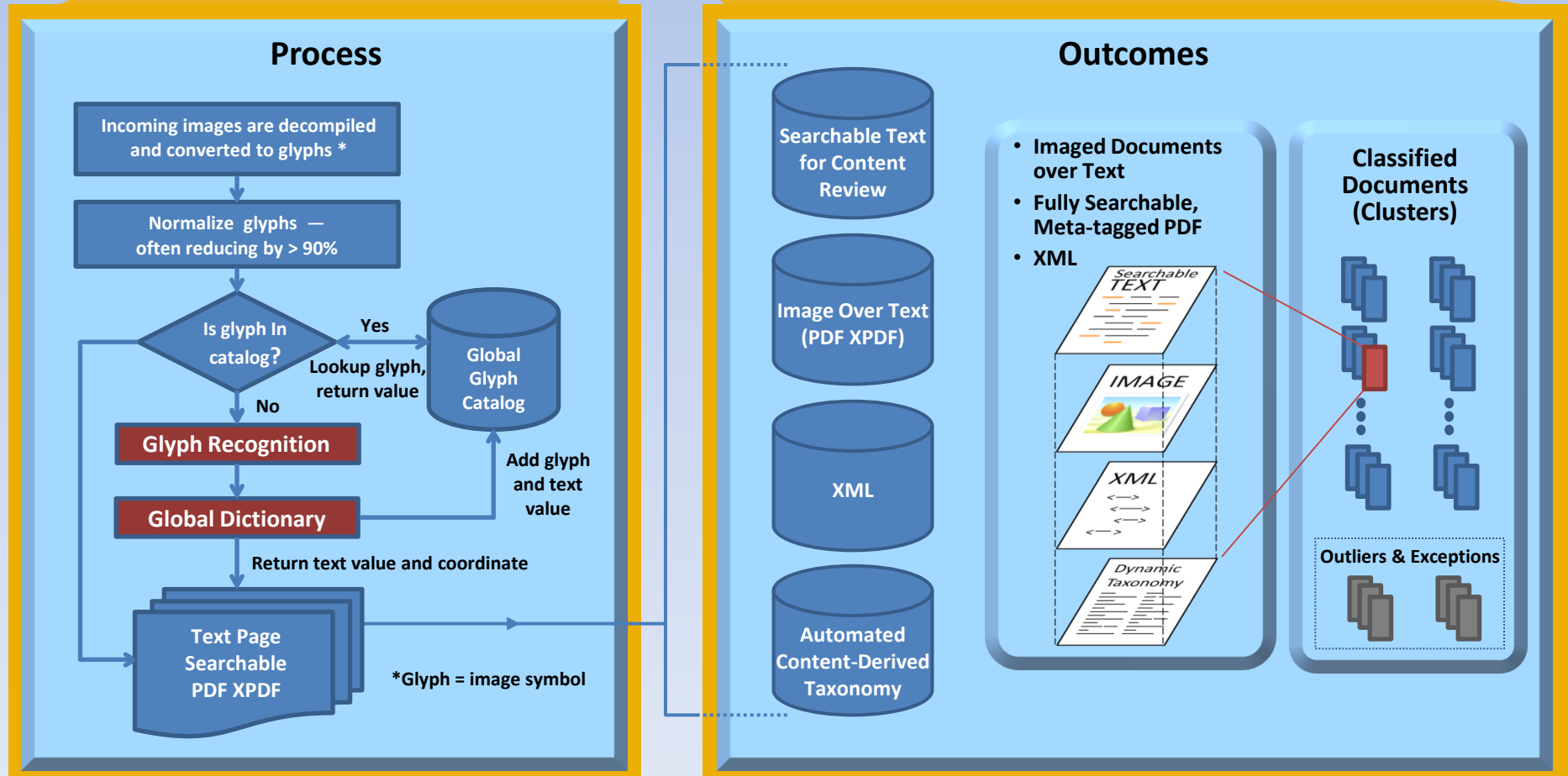
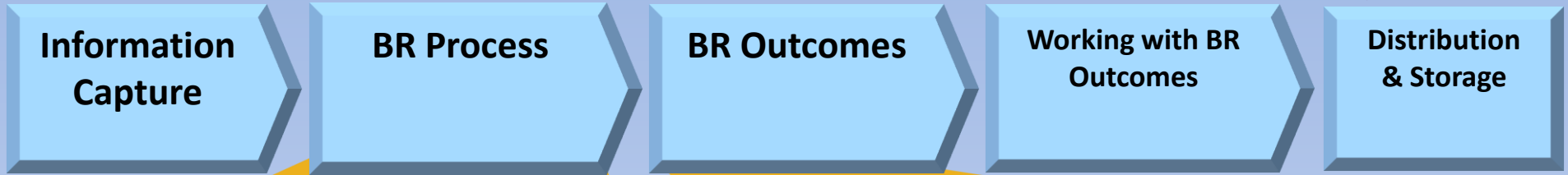


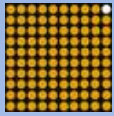
## Image-Based Document Classification Technology

*Technology groups documents together based on visual similarity, not textual analysis, permitting rapid development of classification-specific, location-based data extraction rules.*

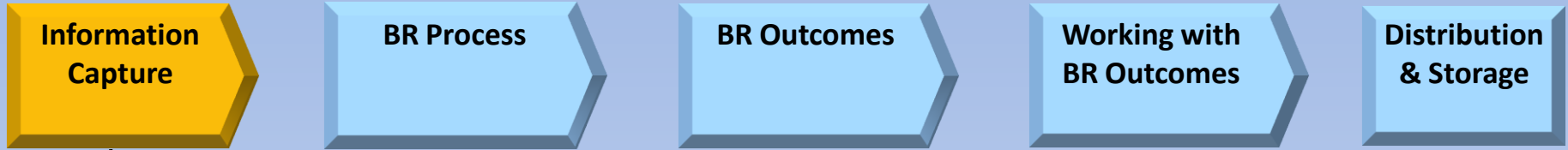


# BeyondRecognition Process



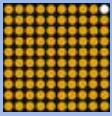


# BeyondRecognition Information Capture

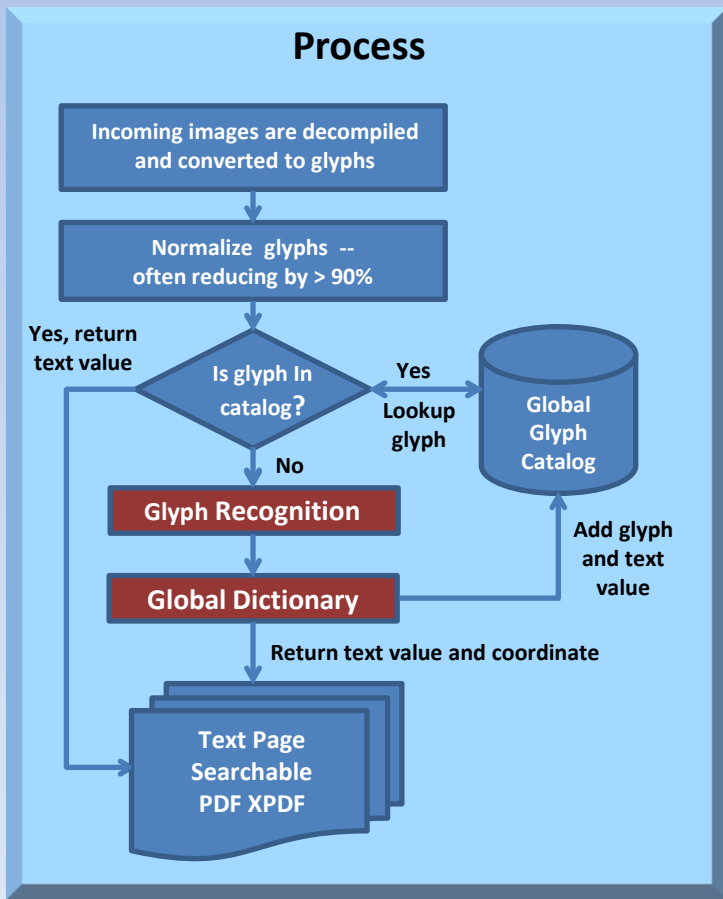
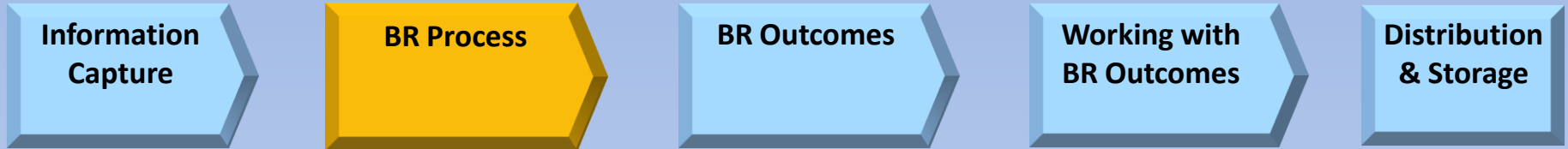


## Information Capture

- BR is compatible with the **widest variety** of document and information capture streams:
  - ✓ Bulk/Batch/Archive
  - ✓ In-house/Outsource
  - ✓ Scanned or Native Format
  - ✓ Normal Line-Of-Business (LOB) applications:
    - Documentum
    - OpenText
    - SharePoint
    - Alfresco
    - OnBase
    - *And more...*

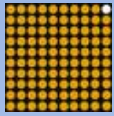


# BeyondRecognition Process

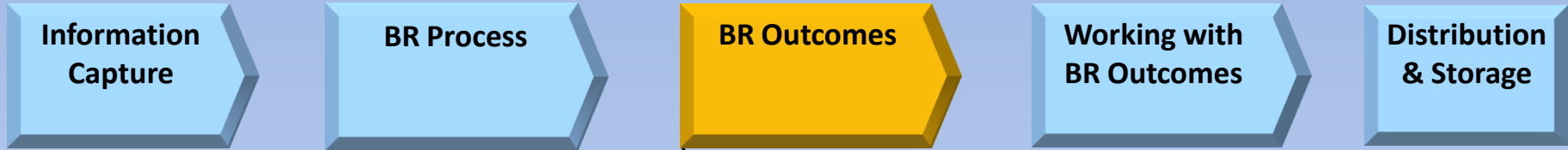


## BeyondRecognition Process

- **Glyph Scraping** – creates an average of 1,500 glyphs (“symbols”) per page, while retaining **Positional Awareness and Context**.
- **Automatic Visual Recognition** compares glyphs to global catalog typically containing hundreds of thousands of glyphs or more; glyphs match regardless of spatial position, rotation and scale.
  - Catalog automatically updated to recognize new glyphs.
  - Text conversion verified against most representative glyph for the specific document cluster. Use of **Glyph Catalog** and **Global Dictionary** typically produces word accuracy > 99.5%.
- BR can automatically perform **Logical Document Boundary Determination**
- **Cascading Auto Correction** of exceptions can automatically update character corrections in millions of documents in a single operation.
- **Automatic Classification** of documents into clusters based on **Visual Similarity** – analyzes scanned image or native electronic such as Excel, Word, PDF, CAD and hundreds of other formats – BR doesn’t get “confused” by file types or source.



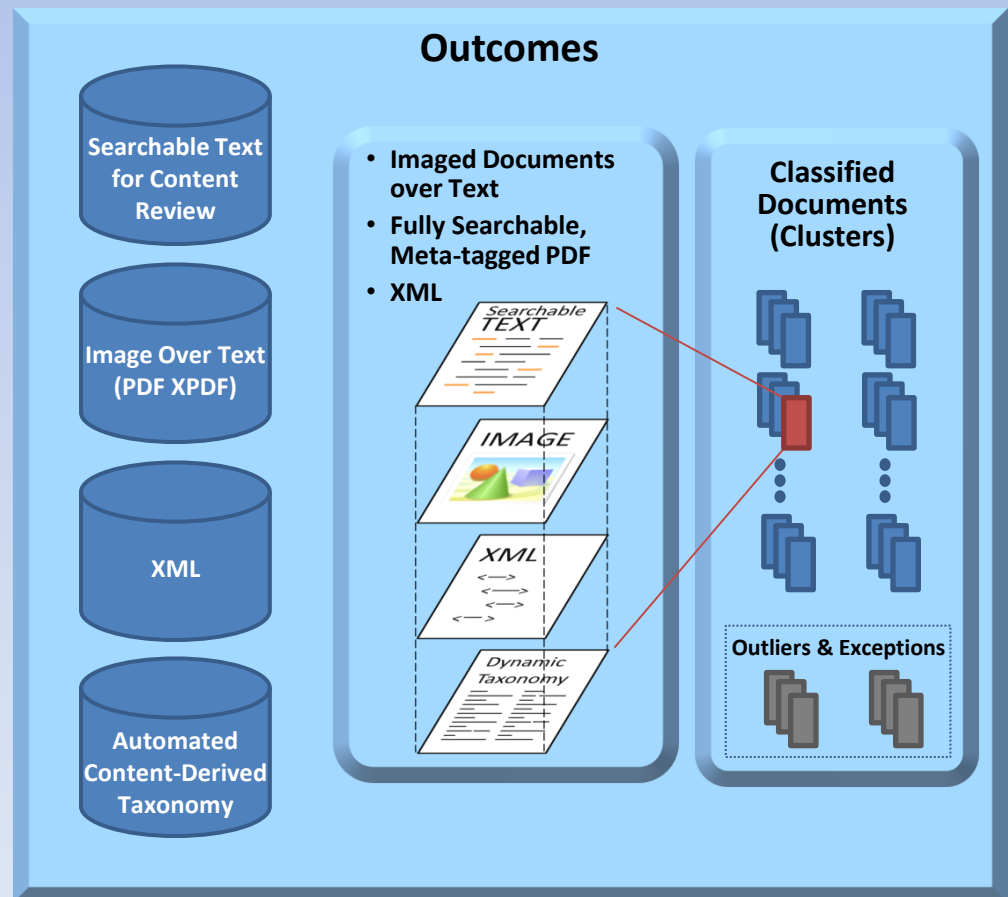
# BeyondRecognition Outcomes

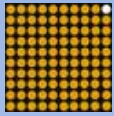


## BeyondRecognition Outcomes

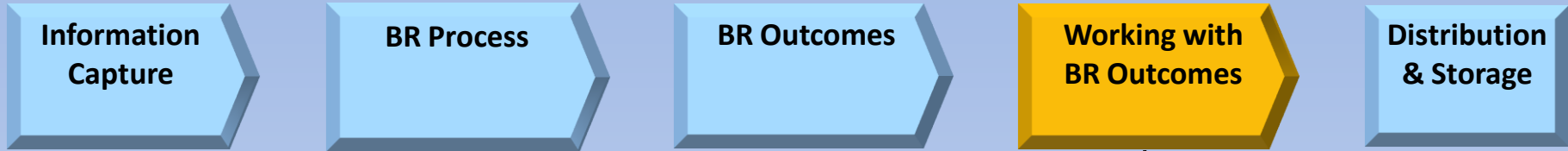
- (1) Clusters of documents as classified
- (2) Searchable text over image PDF/XPDF
- (3) Data extraction directly to XML
- (4) Automatically generated taxonomy

- Documents classified by visual similarity – with **ZERO human intervention or programming**.
- Make retention/reduction/remediation decisions about **millions or billions of documents** by simply looking at the first and last exemplar documents in each document type classification (cluster).
- In most cases, **80% or more** of the total document volume will fall into **less than 2% of the document type clusters** created by BeyondRecognition.
- Readily develop special coding or **data extraction rules for each cluster** via GUI; rules can be position-based (absolute or relative) to page, other glyphs or text – **without creating formal templates**.
- Easily perform **High Volume AutoRedaction** for HIPAA, PCI or other needs using the glyph coordinate location information together with BRs ability to specify patterns, e.g. ### ## ##### for social security numbers.



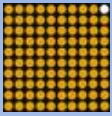


# BeyondRecognition Tools

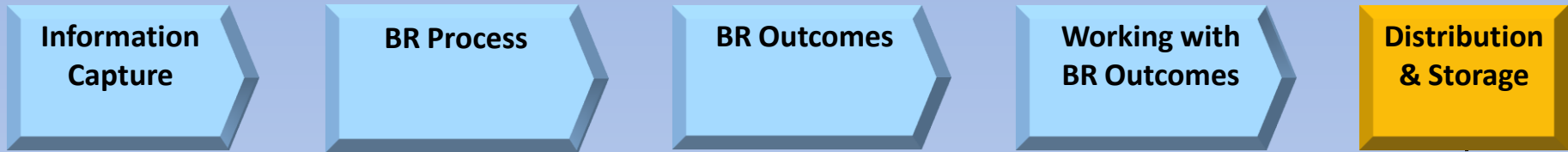


## Working with BR Outcomes

- **Management, Extraction and Review Tools**
  - Visual Review & Drill Down of classified (clustered) documents
  - Extraction of any data on document to XML; easily specify data elements on document image
  - Create fields on the fly
  - Text, Boolean, Symbolic and Positional Search – and any combination
  - Cascading Visual Correction
  - Rules-based Mass Redaction
  - Automatic Content-Derived Taxonomy Creation
- **Application & Rule Development**
  - Customizable Field Extraction
  - Specifying data to extract is as simple as clicking and dragging to draw a box on a document image
  - Easy field creation, including by positional reference
  - C language and .Net based platform, hooks available for custom functionality
- **Quality Control**
  - QC can be done at page level for everything, cluster level, and/or persistent coded level
  - Apps can be readily built to implement quality control procedures – with same or less staffing
- **Overall**
  - Substantial reduction in staffing at all levels – 80-90% in most cases
  - Drastically reduced turn-around time – often from days to minutes or hours



# BeyondRecognition Distribution/Storage



## Load to Storage, Distribute, Leverage

- Load into BeyondRecognition mass storage or 3<sup>rd</sup> Party
- Routing, retention or other actions can be based on visual similarity classification, database, business rules processing, etc.
- Extend the application of “low information density”, unstructured and “intractable” document types to advanced analytics and decision support tools.
- Mitigate risk exposure due to unidentified personal, proprietary or regulated information in documents that are otherwise difficult to classify and process, i.e. scanned paper forms, carbon/carbonless copies, native electronic format engineering drawings, watermarked or poor image quality, time cards, medical film, and industrial film (e.g. as used for ultrasonic, x-ray/gamma imaging and QA documentation of welds and structural components).
- Extend access of information to appropriate stakeholders and processes – from a perspective of **visual content awareness**.

For additional information or to view a  
demonstration contact:

Xpiori, LLC

720-439-9892

[www.xpiori.com](http://www.xpiori.com)

