Growing SharePoint from small libraries to large scale repositories and massive archives

Mirjam van Olst

About Mirjam

Microsoft CERTIFIED Master

MVP Microsoft Most Valuable Professional

Macaw

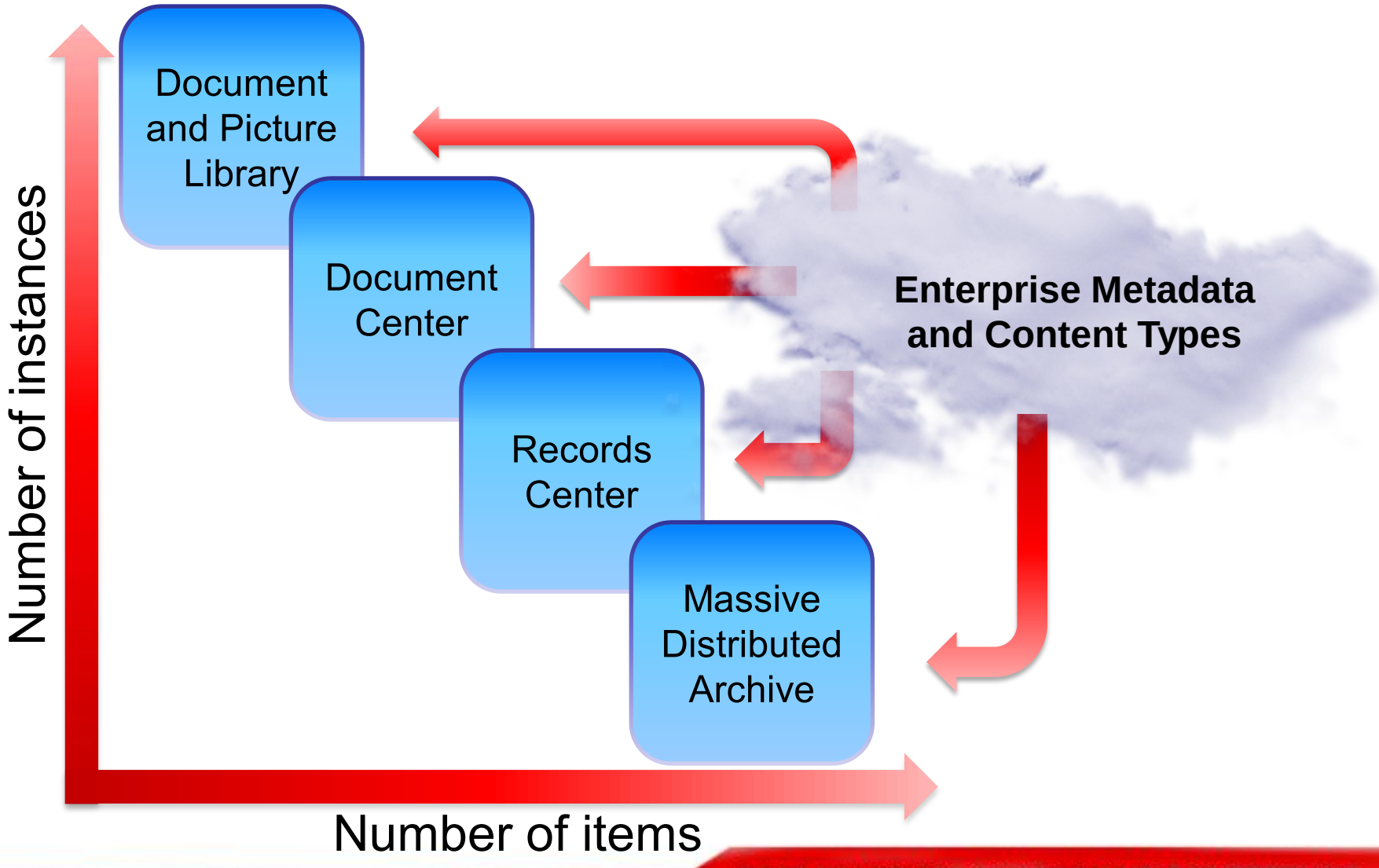
DIWUG

Blog: http://sharepointchick.com

Email: mirjam@macaw.nl

Twitter: @mirjamvanolst

List Scalability



Document and Picture Libraries

- Average size: up to 500 docs
- Used extensively by projects and teams
- Created ad hoc throughout the enterprise
- Libraries are not managed
- Use of content types and site columns for reusability

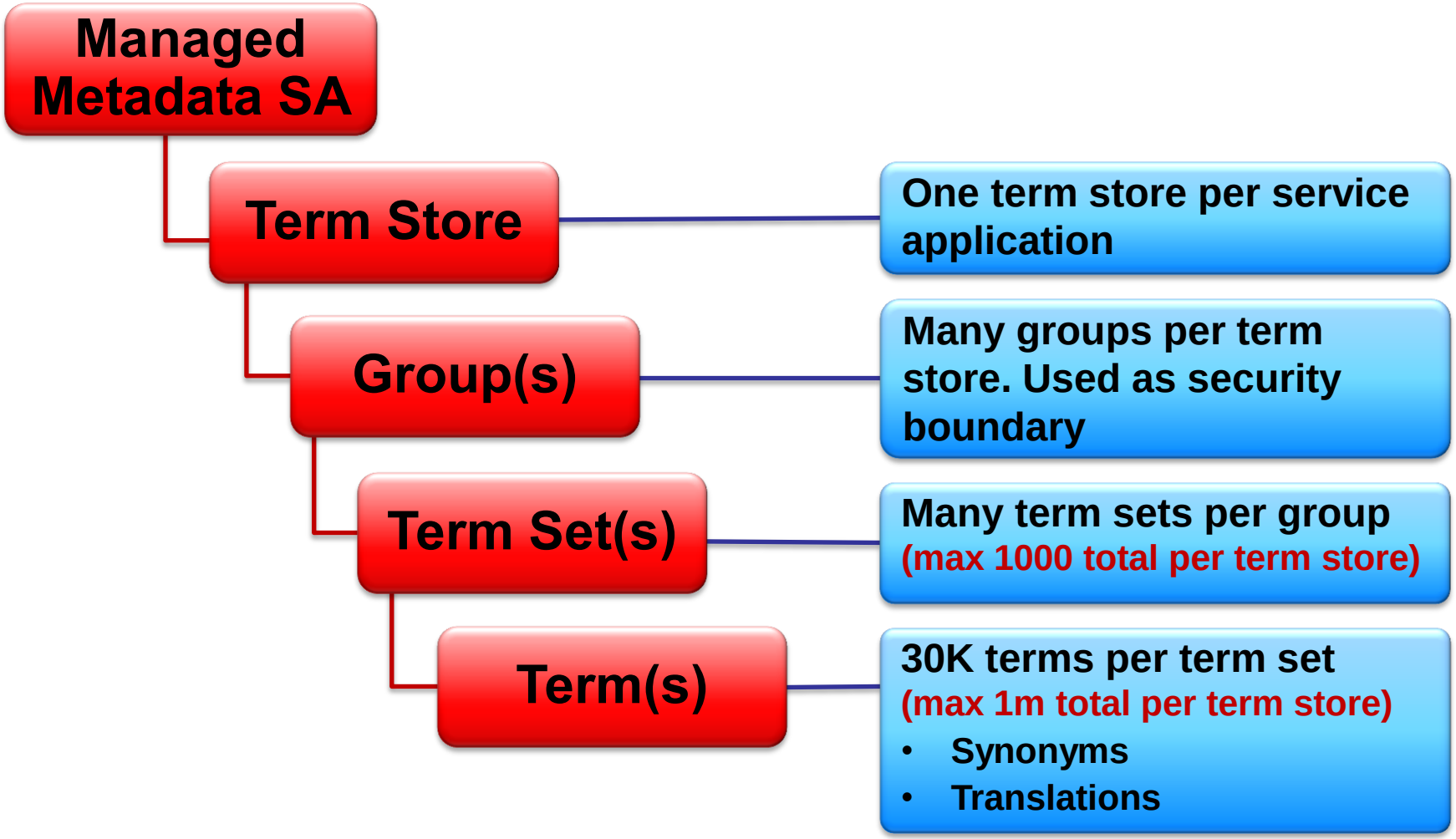
Document and Picture Libraries

- Examples
  - Library for storing a management team's work in progress docs
  - A library spun up for collaborating within a particular project
  - A library on your my site for sharing docs with colleagues
  - A picture library to share photos taken at a company dinner

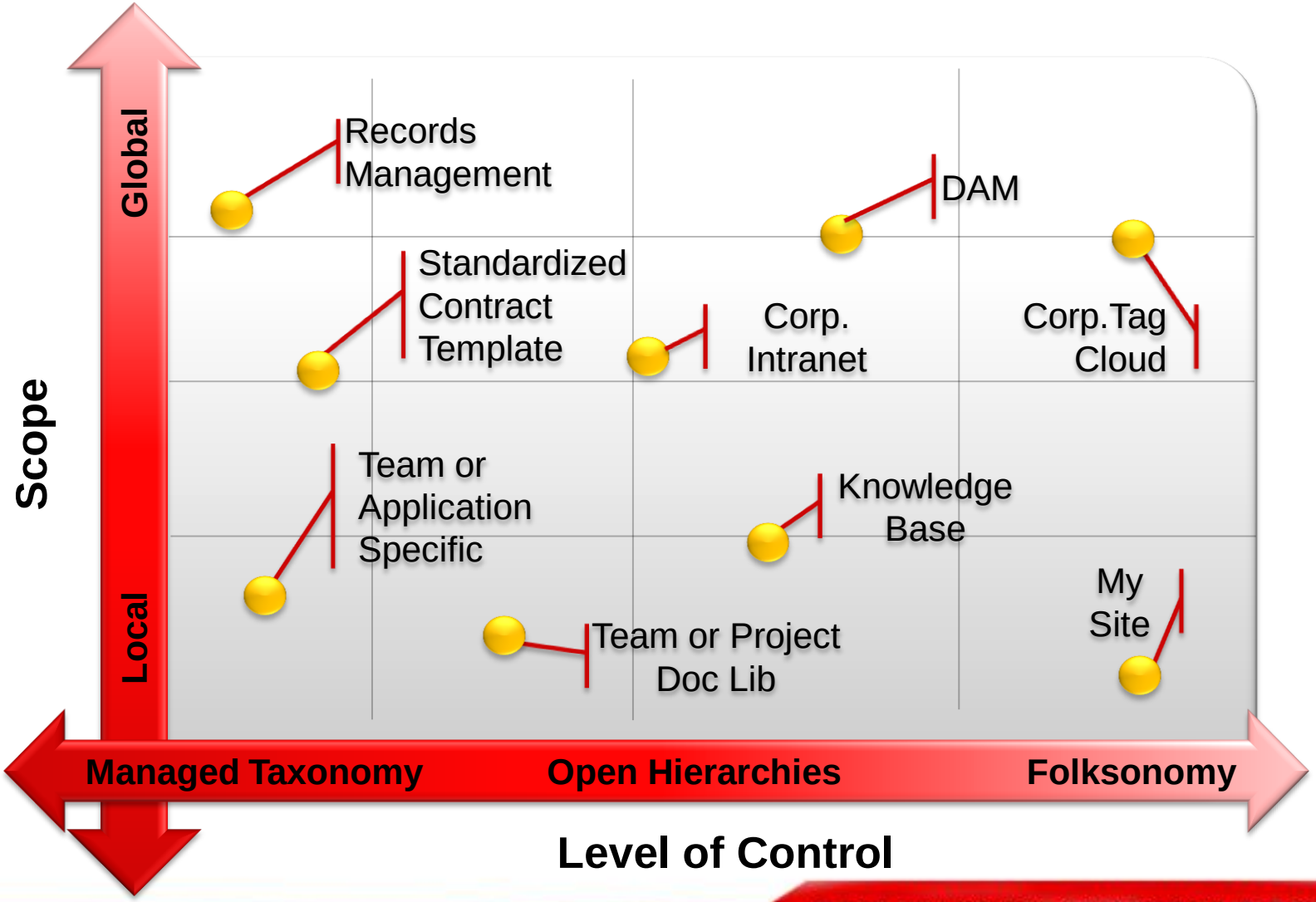
Demo

DOCUMENT LIBRARIES

Term Store


Managed Metadata SA
Term Store
One term store per service application
Group(s)
Many groups per term store. Used as security boundary
Term Set(s)
Many term sets per group
(max 1000 total per term store)
Term(s)
30K terms per term set
(max 1m total per term store)
• Synonyms
• Translations

Term Scope and Control



Best Practices

- Keep your document libraries small
- Make sure documents get archived and libraries get cleaned up
- Make using the term store easy and intuitive for end users
- Give users the right amount of control on the right term sets

Worst Practices

- Making it too difficult for end users
  - To create libraries
  - To use libraries
  - To use metadata
- Users will simply go back to using network shares, or to uploading documents without any metadata

Document Centers

- Average size: 500 – 500.000 docs
- Managed by one or a few subject owners
- Active documents that are being authored and consumed
- Structured navigation needed to help users find what they are looking for
- Correct use of metadata important for findability

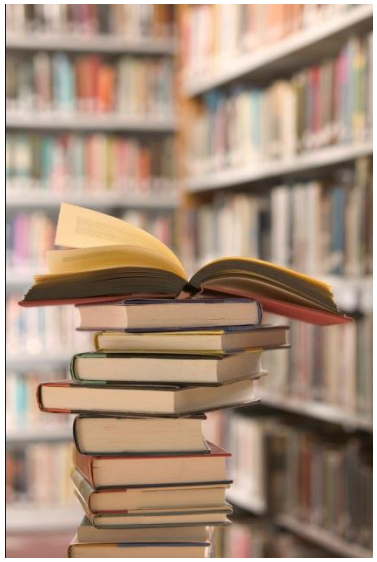
Document Centers

- Examples
  - RFP library for a sales force
  - Contracts library
  - Spec library for an engineering team
  - Brand images repository for marketing
  - HR policies and procedures

Document ID’s

- Unique within a site collection
- Custom prefix
- Not on by default
- Not for list items, docs only
- Kept during operations, except for copy
- Assigns “static url”
- Adds to Document & Document Set Content Types

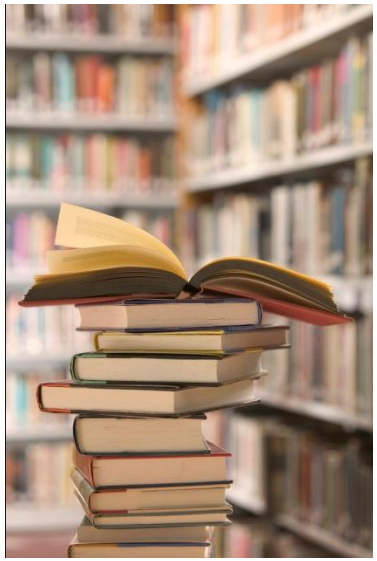
Metadata Navigation



- Navigation Hierarchies
  - Filter on a metadata hierarchy like folders, content types, choice fields, or managed metadata term sets
  - Container contents will be filtered on selected item from the hierarchy
  - Only one filter at a time

Metadata Navigation

- Key Filters
  - Can filter on a broad range of field types
  - Any number of key filters can be applied in combination with a selected navigation hierarchy.
  - “All Tags” queries any managed metadata column
  - From the root folder items from any folder will be returned



List throttling

- Sets limits on how severely users can put the beat down on your servers
- Limits the amount of rows that can be retrieved for a list or library at any one time

List throttling

- Examples of when list throttling will kick in:
  - Viewing data in a list or library
  - Sorting a large list on a non-indexed column
  - Deleting a web with large flat lists
  - When developing solutions while using for instance SPList.Items (depending on settings)

List throttling - options

- In Web Application general settings:
  - List View Threshold –maximum number of items that can be retrieved in one request. Default: 5,000. Smallest: 2,000.
  - Object Model Override –to enable retrieving items through the object model, up to the List view threshold for auditors and administrators
  - List View Threshold for Auditors and Administrators –only relevant when Object Model Override is yes and only via the object model
  - List View Lookup Threshold – only relevant for lookups
  - Daily Time Window for Large Queries –block of time during the day when list throttling limits are not enforced

List throttling - exceptions

- Scenario 1:
  - Box administrator on the SharePoint web server
  - At least Read rights to the list data
- Scenario 2:
  - Using object model SPList class
  - SPList.EnableThrottling = False

Demo

DOCUMENT CENTERS

Best Practices

- Make sure your document center gets managed
- Plan for document retention
- Decide on standard document id provider vs a custom one
- Use managed metadata navigation and list throttling to keep large libraries manageable

Worst Practices

- Using the same prefix for document ids in more than one site collection
- Putting up the throttle limit significantly
- Having people browse your environment as a local admin on the server
- Retrieving more items than necessary when developing SharePoint solutions

Records Centers

- Average size: millions to tens of millions of docs
- Managed by a dedicated team of content stewards
- Users hand off their documents to the archive
- Document are finished, read-only and meant for broad consumption
- Virtual folders will help users to find what they are looking for

Records Centers

- Examples
  - Corporate records archive
  - Knowledge management repository
  - Centralized best practices repository

Recordization in 2010

In Place

- Works best in well established sites (e.g. knowledge management repositories)
- Records are actively used and sit alongside non-records
- Requires tight integration between RM team, IT, and content stewards

Archive

- Centralized vault ingests less managed content
- More traditional; easier to implement hierarchal file plan
- Provides one centralized view and location for all records across enterprise

Content Type Syndication

- Synchronizes content types across site collections
- Select site collection as content type hub in central administration
- Set managed metadata service proxy to consume content types from content type store
- Turn on publishing for content types in hub
- Run timer jobs to make content types available in other sites

Demo

RECORDS CENTERS

Best Practices

- Carefully plan document retention periods for all document types
- Have a separate site collection to serve as a managed content type hub

Worst Practice

- Use both in-place record declaration and archiving of records without a plan
- Insisting on using a single records center, even though your company has to much data or too many documents to manage in a single site collection

Massive distributed archives

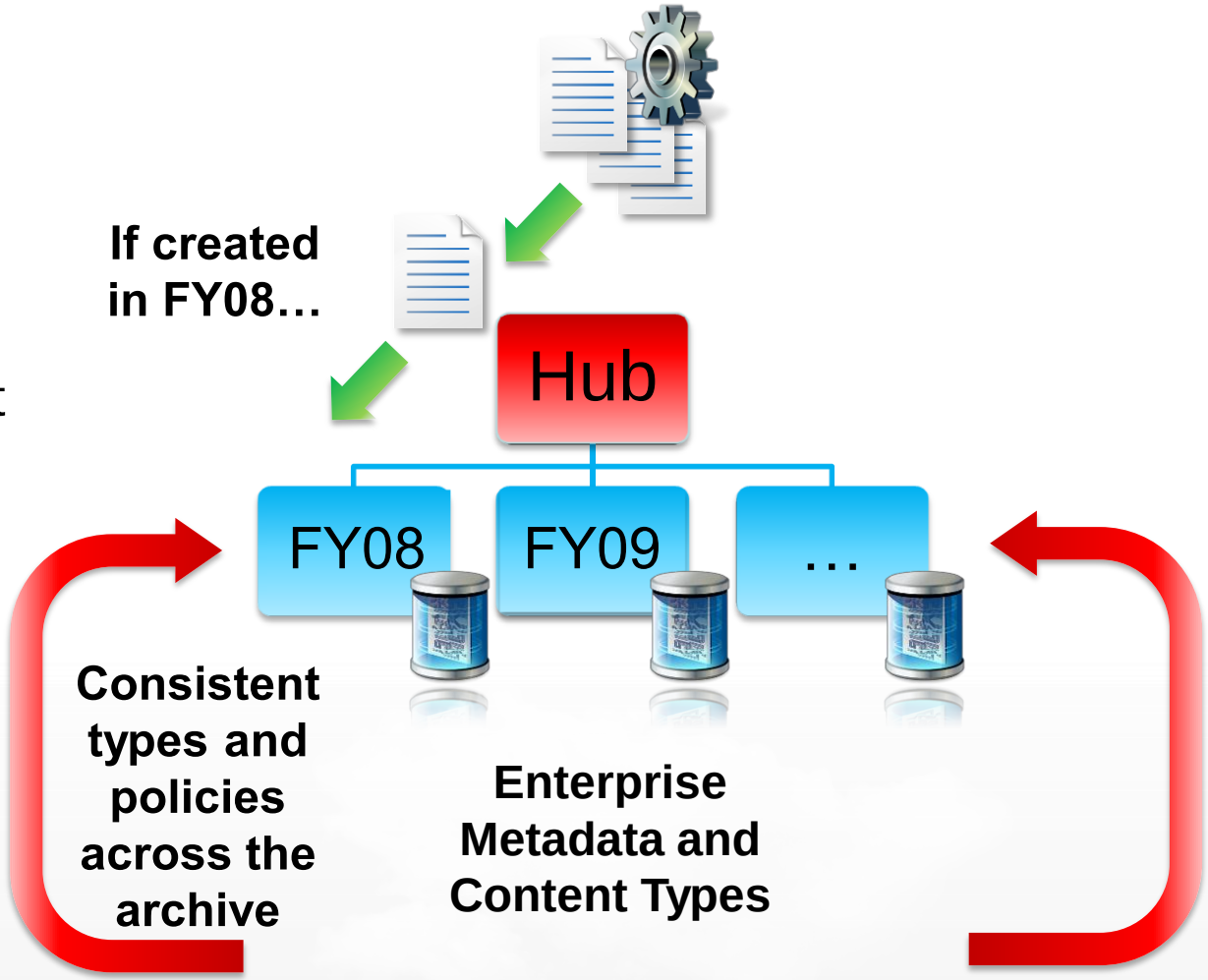
- Key characteristics
  - Average size: hundreds of millions of docs
  - Managed by a dedicated team for the archive
  - Content usually gets added by automated processes
  - Lots of process automation to deal with high volume
  - Backend systems, not users, are primary submitter
  - Logical organization and hierarchy is key

Massive distributed archives

- Examples
  - Archive for a large government agency
  - Yearly archive of insurance forms

Massive distributed archives

- Scale is achieved with a distributed architecture
- Content organizer can route content to correct site collection in the archive
- Content type syndication enables central management of distributed archive
- FAST search is used to retrieve content



Content Organizer

- Automatically routes documents to different libraries and folders
- Can be used to automatically create new folders after a certain # documents have been added
- No user intervention required
- Does not work for lists, just libraries
- DropOff folder created when “Content Organizer” site feature is enabled

Content Organizer Settings

- Redirect Users to the Drop Off Library
- Sending to another site enables documents to be redirected to other sites, or site collections
- Folder Partitioning to automatically distribute documents across folders
- Duplicate Submissions to add new versions, or add unique characters to the filename
- Preserving context to keep audit logs and properties

Content Organizer Rules

- Rules primarily based on content types
- Conditions can be added based on document properties
- Target Location can either be a library in the current site, or a different site or site collection
- Target location sites need to:
  - Have the content organizer feature activated
  - Be registered in Central Administration
  - Have the content type available

Demo

MASSIVE DISTRIBUTED ARCHIVE & CONTENT ORGANIZER

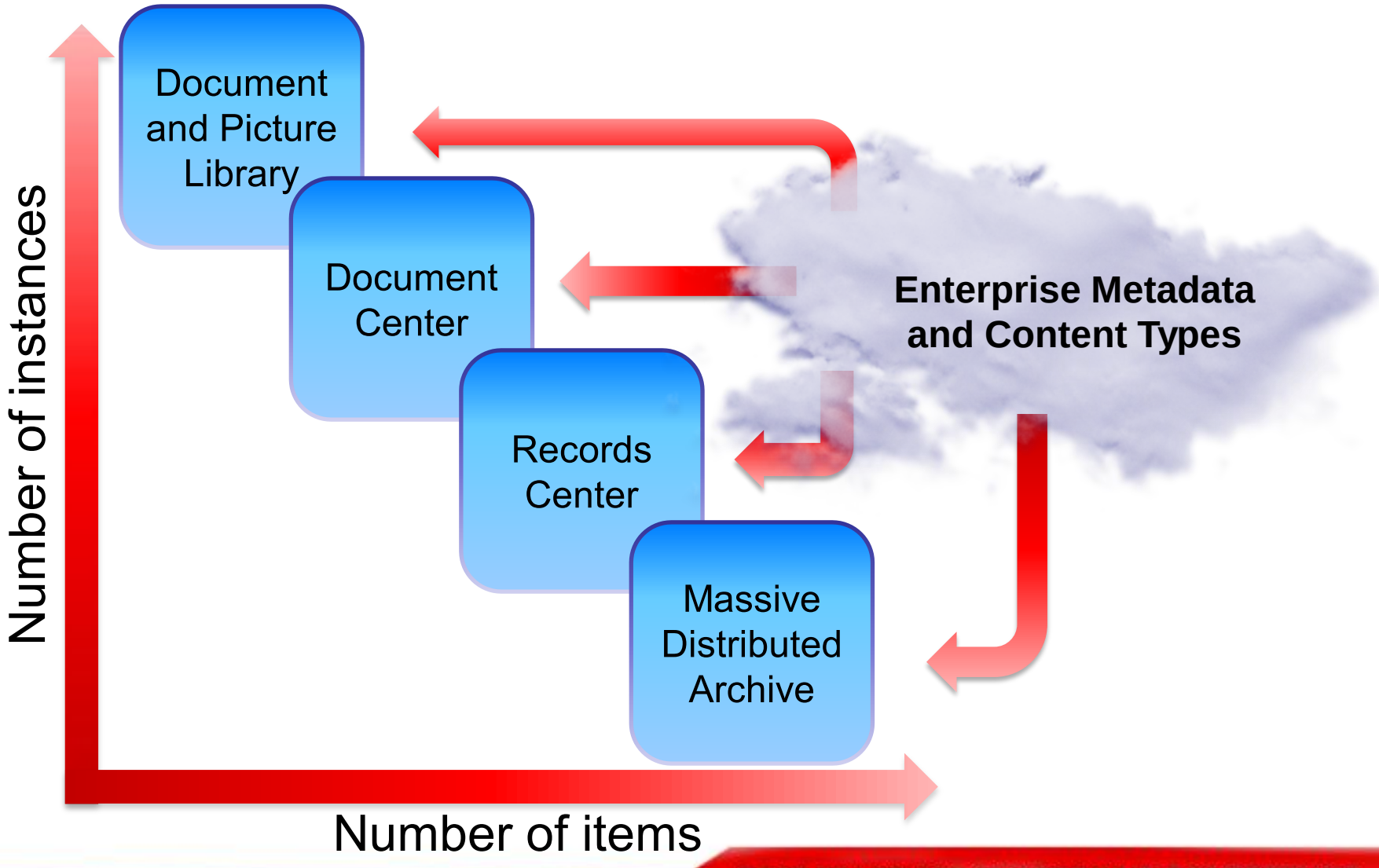
Best Practices

- Create more than one records center if your company has a lot of data
- Organize data over multiple records centers based on metadata that makes sense to your organization
- Use FAST search to enable users to find documents from the archives

Worst Practices

- Use more than one records center because it’s a cool thing to brag about

List Scalability



Growing SharePoint from small libraries to large scale repositories and massive archives

Please be sure to fill out your session evaluation!