

Cataloging for the Preaching and Worship Portal

Harry Plantinga

April 10, 2014

The Preaching and Worship Portal (PWP) will provide a portal for pastors and worship leaders to find preaching and worship resources, so it will need to have a large selection of such resources indexed. Resources such as sermon starters, blog posts and essays, images, and video will be reviewed and tagged by human catalogers, assisted as much as possible by programs that automate parts of the process.

A web crawler will visit designated websites and locate all the pages at those sites, putting them into a cataloging queue for human cataloging. It will fill in as much information about the resources as possible. Then a human will catalog the resource, checking the computer-generated entries and filling out missing information on a cataloging form. When that is complete, the resource may be made available for search by clicking a “publish” option.

Changes may be made to the cataloging record after it has been published. For example, a person writing resource recommendations might look at a list of newly added resources, browse a few, and choose to add recommendations to some, feature them, or otherwise edit the cataloging data.

The crawler will visit partner sites regularly to find new Web pages and other resources, which it will place in a cataloging queue. In addition, there should be a crawler that visits all resources in the database to check that they are still available.

A person indexing 15 resources per hour could index 10,000 resources working half time for eight months. This would not include writing descriptions or recommendations. That number of resources would be about 40 per lectionary week or 8 per chapter of the bible—a decent start.

Cataloging Resources

One primary value of the PWP will be its ability to efficiently find resources meeting the needs of preachers and worship leaders. The kinds of queries we hope to support are identified elsewhere.¹ In order to support these queries, we will need to gather various pieces of information for each document. Here is a first draft of the information that should be added:

1. Title
2. URL
3. Teaser
4. Author/source(s)
5. Principal scripture verses
6. Named recommendation, source (s)
7. Lectionary week

¹ See “Semantic Search for the Preaching and Worship Portal,” CCEL TR #9

8. Tags (Entity tags, resource types, liturgical elements, season, holiday)
9. Published, featured

Teaser: the teaser is a short excerpt or description of about 24 (or 20-30) words, which will be shown in search result lists and in other circumstances to give an indication of the contents of the resource. See, for example, the teasers on Google search results pages or the Arts and Letters Daily home page. We can provide an automatic suggestion consisting of either the meta description of the resource provided by its author or the first 24 words. For non-textual resources such as images and video, this should be a brief description.

Author/source(s): This will be a pointer to an author or source entry in the database. It may be repeated if there are multiple sources. If there is no existing record for the author or source, the cataloger will be asked to create one. It should have a name, affiliation, description, denomination, and an image for the source.

Principal scripture verses: The principal scripture verses box can have a list of verses that the resource is principally about. For example, a sermon starter for a sermon on Matthew 5:8 would list that verse. A commentary on Matthew 5 would list the whole chapter. We will also support searching by scripture passages merely mentioned, but the crawler should find these automatically. The scripture passages may be entered as they are commonly written, but internally they will be stored as a list of integer verseIDs.

Named recommendations: This section enables editors to add brief (typically one sentence) recommendations of resources. These will appear along with the name of the recommender on a description of that resource. This field is repeatable.

Lectionary week: This enables the cataloger to indicate that this resource is appropriate for one or more particular lectionary weeks. There should be a box to enter a reason for each recommendation, typically a thematic match or a match to one of the scripture passages. The lectionary week selector should appear as a pop-up showing all the lectionary weeks and any themes for the weeks, with a checkbox and reason box by each week. Lectionary weeks containing scripture passages matching the “principal scripture verses” field should be pre-checked and the reason (“scripture passage match”, listing scripture passage) filled out. When the pop-up is closed, the list of lectionary weeks and reasons should be displayed textually. Note that there are denominational differences in the common lectionary; we should support all denominations.

The *Tags* will be added by checking checkboxes beside a list of possible tags, resource types, liturgical elements, and the like. A sample page for cataloging, showing a draft list of tags and other categorizations, is appended. Internally, these will be stored as a list of integer values (entityNIDs) indicating the entities with which the resource is tagged.

Published, Featured: These Boolean fields will be represented as checkboxes. Checking the “published” box makes it eligible to appear in search results. The “featured” checkbox should be checked to make this resource eligible to appear in a list of featured resources on the home page.

Support for Cataloging

We will need to create a Web page that shows resources queued for cataloging. When one is clicked, a cataloging form for the resource should appear, showing a form along the lines of the example below. Additionally, the resource itself should be opened in a separate window. When the cataloging form is filled out and “Submit” clicked, if the “publish” checkbox is selected, the resource should be marked as published and made available as a search result. It should also be removed from the cataloging queue. (It should not be re-added if the resource is later unpublished.) The cataloger should also be able to click a “delete” button to remove the entry from the queue without cataloging it, and the crawler should not later re-add the resource to the queue.

There should be an “edit” tab or link beside a view of a resource that appears for editors. Clicking it should open the cataloging form for the resource. There should be a log of the date and time when the resource was originally cataloged and when it was edited, with the name of the person performing the task. These should appear on the cataloging form for the resource.

There should also be a page showing the resources that have been recently added, and another showing the list of resources that are unpublished but not in the cataloging queue. These could conceivably be tabs on the page showing documents queued for cataloging. These lists should be searchable.

There should be a statistics page showing statistics about resources for a given date range (including a calendar for selecting the date range and one-click options for yesterday, last week, last month, and all time). It should show the number of resources of the various types, the number cataloged or edited by each editor, the number in the queue, the total number of resource views, and the top 10 for the given date range.

Web Crawler

The Web crawler visits websites listed in a box on a configuration Web page and adds all previously unseen resources to the cataloging queue. It should create cataloging records for the resources, filling out as much information as possible, and marking them as unpublished. It should fill out the URL, and it can use the document title as the title. For the teaser, it can use the document meta description, if available, or the first 24 words of the body of the document.

It should parse the document, looking for scripture passages, and put the verse IDs in the “Additional scripture passages” field. If it finds a passage in a header element near the top of the document, it can add that passage to the “Primary scripture passages” field.

It may be possible to guess at the tags that should be clicked through linguistic analysis. Initially, the crawler could simply look for the keywords used to indicate entities and check entities when one of their keywords is present. Experience should tell us whether these guesses are helpful or whether they can be improved. Later, a more sophisticated analysis such as keyword extraction may be worth exploring.

The crawler should identify itself as the “PWP Crawler”, and it should respect robots.txt files. There should be a Web page for the crawler giving information about it and a way to contact us if there are problems.

There should also be a program that regularly visits all the resources in the database to make sure that they are still available. If they are unavailable for two visits, they should be marked thus in the database and not shown as search results. Occasionally it can recheck resources marked “no longer available” and make them available again if they reappear.

There should be a Web page controlling the crawler—allowing an editor to specify sites to be crawled and the number of days between crawls for each site. It should also give statistics such as the number of sites crawled in the last day and week, the number of resources added to the queue, and the like.

PWP Resource Cataloging

Title: _____

URL: _____

Teaser (20-30 word excerpt or description):

Source [Select list: select an available source, or add a new source]

Resource type: [Select list: Sermon Illustration, Sermon Starter/Outline, Complete Sermon, Commentary/exegesis, Blog/essay, Reflection questions, Book citation, movie citation, Topical guide, Map, Image, Video, Hymn/song, Drama, Activity, Children's resource, Complete liturgy, Liturgical element]

Element of worship: [Select list: Agnus Dei, Amen, Benediction, Call to worship, Close of worship, Communion, Confession, Credo, Gloria, Gospel acclamation, Illumination, Kyrie, Lord's Prayer, Memorial acclamation, Offertory, Praise, Prayer, Response, Sanctus]

Scripture verses this resource treats or concerns: _____

Recommendation: _____ Source: _____ (repeatable)

Events

- Baptism
- Birth
- Death
- Divorce
- Funeral
- Wedding

Special Days

- All saints' day
- Ascension
- Ash Wednesday
- Christmas
- Easter
- Epiphany
- Good Friday
- Pentecost
- Reformation day
- Transfiguration

Church Seasons

- Advent
- Christmas
- Easter
- Epiphany
- Lent
- Pentecost

Theme clusters

- Beatitudes
- Fruit of the Spirit
- Gifts of the Spirit
- Lord's Prayer
- Parables
- Sermon on the Mount
- Seven deadly sins
- Ten Commandments

Christian life

- Beauty
- Conversion
- Culture
- Family
- Fear
- Grief
- Guilt
- Healing
- Leadership
- Marriage
- Ministry
- Money
- Penance
- Pilgrimage
- Politics
- Poverty
- Promises
- Religion
- Repentance
- Reverence
- Sexuality
- Spirituality, piety, devotion
- Suffering
- Temptation
- Tithing
- Tongues
- Work

Spiritual practices

- Confession
- Discipleship
- Evangelism
- Fasting
- Forgiveness
- Praise
- Prayer
- Rejoicing
- Thanksgiving

Virtues

- Compassion
- Contentment
- Faith
- Faithfulness
- Gentleness
- Goodness
- Hope
- Hospitality
- Humility
- Joy
- Kindness
- Love
- Mercy
- Obedience
- Patience
- Peace
- Self-control
- Trust
- Wisdom

Sins, vices

- Addiction
- Anger
- Blasphemy
- Corruption
- Envy
- Folly
- Gluttony
- Gossip
- Greed
- Hatred
- Hypocrisy
- Idolatry
- Lust
- Lying
- Pride
- Selfishness
- Sloth, despondency
- Stealing

Ten Commandments

- Idolatry
- Making graven images
- Taking God's name in vain
- Remember the Sabbath
- Honor your father and mother
- Murder
- Adultery
- Theft
- False witness
- Coveting

Church life

- Children/youth
- Church
- Discipline
- Liturgy
- Revival
- Worship

Trinity

- Trinity
- God the father
- Holy spirit
- Jesus Christ
- Ascension
- Crucifixion
- Incarnation
- Nativity
- Resurrection
- Transfiguration

Attitude

- Irony
- Humorous

Theological topics

- Angels
- Covenant
- Creation
- Election
- Eschatology
- Ethics
- Good works
- Glory
- Grace
- Heaven
- Hell
- Image of God
- Judgment
- Justice
- Kingdom of God
- Law
- Lord's Supper
- Morality
- Miracles
- Providence
- Redemption
- Salvation
- Sabbath
- Sacrifice
- Sanctification, holiness
- Scripture
- Shalom
- Sin, evil

Status

- Published
- Featured

[Submit]

[Delete]