

2015 IFLA News Media Conference

Online news media in Mimer:
the challenge of metadata quality
in e-legal deposit in Sweden /

Joakim Philipson (National Library of Sweden)

This is what we got:



Foto: Fredrik Sandberg/T1

Personbil rammade busskur i Stockholm

Publicerad 12 feb 2015 20:45 [A](#)[A](#)[A](#)

Polisen försökte stoppa bil som körde vårdslöst söder om Stockholm på torsdagskvällen.

Färden slutade med att bilen rammade en busskur i Farsta.

En kvinna i 45-årsåldern befann sig i busskuren och fick flygas till sjukhus i ambulanshelikopter.

Hennes tillstånd ska inte vara livshotande, meddelar polisen under torsdagsnatten.

ANNONS:

This is what we get:

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0">
  <channel xmlns:dcterms="http://purl.org/dc/terms/" xmlns:media="http://search.yahoo.com/mrss/" xmlns:georss="http://www.w3.org/2003/01/geo/wgs84_pos#" xmlns:atom="http://www.w3.org/2005/Atom"
    xmlns:content="http://purl.org/rss/1.0/modules/content/">
    <title>Expressen</title>
    <link>http://www.expressen.se/</link>
    <description>Expressen</description>
    <atom:link type="application/rss+xml" href="http://www.expressen.se/elibrary.xml" rel="self"/>
  <item>
    <guid>http://www.expressen.se/gt/bil-brann--spred-sig-till-en-grillrestaurang/</guid>
    <link>http://www.expressen.se/gt/bil-brann--spred-sig-till-en-grillrestaurang/</link>
    <pubDate>Fri, 13 Mar 2015 00:08:58 +0100</pubDate>
    <dcterms:publisher>http://id.kb.se/organisations/SE5560254525</dcterms:publisher>
    <title>Bil brann - spred sig till en grillrestaurang</title>
    <dcterms:accessRights>gratis</dcterms:accessRights>
    <dcterms:format>text/html</dcterms:format>
    <description><p>En bil och en grillrestaurang brinner på Agnesbergsvägen i Göteborg.</p> <p>- Det är en parkerad bil som börjat brinna och där elden spridit sig till den intilliggande byggnaden, säger Dennis Olsson, larm- och ledningsoperatör på Räddningstjänsten i Göteborg.</p></description>
    <content:encoded><p>Det var vid 23.30-tiden som Räddningstjänsten fick larmet om branden på Agnesbergsvägen 20.</p> <p>- När vi kommer till platsen brinner en bil som vi för släcka först och sedan flytta på. Men då har branden spridit sig till en del av taket och rök fyllt en del av restauranterlokalen, säger Dennis Olsson.</p> <p>Grillrestaurangen var stängd när det börjar brinna och inga personer fanns i huset.</p> <p>Texten uppdateras</p></content:encoded>
    <dcterms:creator>Ingrid Norrman</dcterms:creator>
    <category>GT</category>
  </item>
  <!-- ... -->
  <item>
    <guid>http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/</guid>
    <link>http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/</link>
    <pubDate>Fri, 13 Feb 2015 00:46:52 +0100</pubDate>
    <dcterms:publisher>http://id.kb.se/organisations/SE5560254525</dcterms:publisher>
    <title>Personbil rammade busskur i Stockholm</title>
    <dcterms:accessRights>gratis</dcterms:accessRights>
    <dcterms:format>text/html</dcterms:format>
    <description><p>Polisen försökte stoppa bil som körde vårdslöst söder om Stockholm på torsdagskvällen.</p> <p>Färden slutade med att bilen rammade en busskur i Farsta.</p> <p>En kvinna i 45-årsåldern befann sig i busskuren och fick flygas till sjukhus i ambulanshelikopter.</p> <p>Hennes tillstånd ska inte vara livshotande, meddelar polisen under torsdagsnatten.</p></description>
    <content:encoded><p>Det var vid 20-tiden på torsdagskvällen som en polispatrull observerade en utlandsregistrerad Audi som körde vårdslöst på riksväg 73 söder om Stockholm.</p> <p>Polisen fattade beslut om att följa efter och kontrollera fordonet, som körde i södergående riktning.</p> <p>Bilen svängde av vid Farsta där föraren tappade kontrollen och kraschade in i en busskur.</p> <p>- I samband med det så har den kört på en kvinna, säger Lars Källgren, stationsbefäl på Söderortspolisen.</p> <h2>Skador inte livshotande</h2> <p>Kvinnan är i 50-årsåldern och fördes i ambulanshelikopter till Karolinska universitetssjukhuset med allvarliga skador.</p> <p>- Senaste lägesbilden är att det inte är livshotande och att det är stabilt. Det är väldigt glädjande att det verkar ha gått bra, säger Lars Källgren strax efter midnatt.</p> <p>Föraren till Audin var en man i 40-årsåldern som blåste positivt för rattfylleri på plats och togs med av polis för kroppsbesiktning.</p> <p>Mannen är kriminellt belastad och känd av polis sedan tidigare. Under kvällen greps han misstänkt för en rad brott.</p> <h2>Flera brottsmisstankar</h2> <p>- Han är misstänkt för grovt rattfylleri genom påverkan av alkohol och/eller narkotika, grovt vållande till kroppsskada, grov vårdslöshet i trafik, tillgrepp av fortskaffningsmedel och ringa narkotikabrott genom eget bruk, säger Lars Källgren.</p> <p>Polisen har under kvällen hört ett flertal vittnen till händelsen.</p> <p>Under natten till fredag arbetar man med att fastställa uppgifter kring den stulna utlandsregistrerade bilen. Fordonet har tagits i beslag och ska även genomgå en teknisk undersökning.</p></content:encoded>
    <dcterms:creator>Kim Malmgren</dcterms:creator>
    <category>Nyheter</category>
  </item>
  <media:content type="image/jpeg" url="http://y.cdn-expressen.se/images/7a/cb/7acb8b54cabe4ff8bc4c04c4b9dbcb21/16x6/540@70.jpg">
    <media:description type="plain"/>
    <media:copyright>Fredrik Sandberg/TT</media:copyright>
  </media:content>
</item>
</channel>
</rss>
```

... or, after some work, rather this:

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0">
- <channel xmlns:dcterms="http://purl.org/dc/terms/" xmlns:media="http://search.yahoo.com/mrss/" xmlns:georss="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:atom="http://www.w3.org/2005/Atom" xmlns:content="http://purl.org/rss/1.0/modules/content/">
  <title>Expressen</title>
  <link>http://www.expressen.se/</link>
  <description>Expressen</description>
  <atom:link type="application/rss+xml" rel="self" href="http://www.expressen.se/elibrary.xml"/>
- <item>
  <guid>http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/</guid>
  <link>http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/</link>
  <pubDate>Fri, 13 Feb 2015 00:46:52 +0100</pubDate>
  <dcterms:publisher>http://id.kb.se/organisations/SE5560254525</dcterms:publisher>
  <title>Personbil rammade busskur i Stockholm</title>
  <dcterms:accessRights>gratis</dcterms:accessRights>
  <dcterms:format>text/html</dcterms:format>
  <description><p>Polisen försökte stoppa bil som körde vårdslöst söder om Stockholm på torsdagskvällen.</p> <p>Färden slutade med att bilen rammade en busskur i Farsta.</p>
  <p>En kvinna i 45-årsåldern befann sig i busskuren och fick flygas till sjukhus i ambulanshelikopter.</p> <p>Hennes tillstånd ska inte vara livshotande, meddelar polisen under
  torsdagsnatten.</p></description>
  <content:encoded><p>Det var vid 20-tiden på torsdagskvällen som en polispatrull observerade en utlandsregistrerad Audi som körde vårdslöst på riksväg 73 söder om Stockholm.</p>
  <p>Polisen fattade beslut om att följa efter och kontrollera fordonet, som körde i södergående riktning.</p> <p>Bilen svängde av vid Farsta där föraren tappade kontrollen och
  kraschade in i en busskur.</p> <p>- I samband med det så har den kört på en kvinna, säger Lars Källgren, stationsbefäl på Söderortspolisen.</p> <h2>Skador inte
  livshotande</h2> <p>Kvinnan är i 50-årsåldern och fördes i ambulanshelikopter till Karolinska universitetssjukhuset med allvarliga skador.</p> <p>- Senaste lägesbilden är att det
  inte är livshotande och att det är stabilt. Det är väldigt glädjande att det verkar ha gått bra, säger Lars Källgren strax efter midnatt.</p> <p>Föraren till Audin var en man i 40-
  årsåldern som blåste positivt för rattfylleri på plats och togs med av polis för kroppsbesiktning.</p> <p>Mannen är kriminellt belastad och känd av polis sedan tidigare. Under
  kvällen greps han misstänkt för en rad brott.</p> <h2>Flera brottsmisstankar</h2> <p>- Han är misstänkt för grovt rattfylleri genom påverkan av alkohol och/eller narkotika,
  grovt vållande till kroppsskada, grov vårdslöshet i trafik, tillgrepp av fortskaffningsmedel och ringa narkotikabrott genom eget bruk, säger Lars Källgren.</p> <p>Polisen har under
  kvällen hört ett flertal vittnen till händelsen.</p> <p>Under natten till fredag arbetar man med att fastställa uppgifter kring den stulna utlandsregistrerade bilen. Fordonet har tagits
  i beslag och ska även genomgå en teknisk undersökning.</p></content:encoded>
  <dcterms:creator>Kim Malmgren</dcterms:creator>
  <category>Nyheter</category>
- <media:content type="image/jpeg" url="http://y.cdn-expressen.se/images/7a/cb/7acb8b54cabe4ff8bc4c04c4b9dbcb21/16x6/540@70.jpg">
  <media:description type="plain"/>
  <media:copyright>Fredrik Sandberg/TT</media:copyright>
  </media:content>
  </item>
</channel>
</rss>
```

This is (only part of) what we need: ... for storage and preservation

```
<?xml version="1.0"?>
<mets:mets xsi:schemaLocation="http://www.loc.gov/METS/ http://www.kb.se/namespace/mets/kbse_mets_002.xsd http://www.loc.gov/mods/v3
http://www.kb.se/namespace/mods/mods.xsd info:lc/xmlns/premis-v2 http://www.kb.se/namespace/premis/v2/premis.xsd http://www.w3.org/1999/xlink
http://www.kb.se/namespace/mets/SWEIP_METS_xlink.xsd" TYPE="AIP" PROFILE="http://www.kb.se/namespace/mets/kbse_mets_profile_002.xml" OBJID="urn:nbn:se:kb:dark-package-
instance-1406084" LABEL="Personbil rammade busskur i Stockholm" ID="aip.mets.metadata" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:sh="http://www.kb.se/mimer/normalization/kbmets/shared" xmlns:premis="info:lc/xmlns/premis-v2"
xmlns:mods="http://www.loc.gov/mods/v3" xmlns:mets="http://www.loc.gov/METS/" xmlns:marc="http://www.loc.gov/MARC21/slim"
xmlns:cpp="http://www.kb.se/marc/channelrecord/parser">
- <mets:metsHdr RECORDSTATUS="VERSION" CREATEDATE="2015-02-13T11:29:05.472+01:00">
- <mets:agent TYPE="ORGANIZATION" ROLE="ARCHIVIST">
  <mets:name>Kungl. biblioteket</mets:name>
  <mets:note>http://id.kb.se/organisations/SE2021001710</mets:note>
</mets:agent>
- <mets:agent TYPE="ORGANIZATION" ROLE="CREATOR">
  <mets:name>Kungl. biblioteket</mets:name>
  <mets:note>http://id.kb.se/organisations/SE2021001710</mets:note>
</mets:agent>
<mets:altRecordID TYPE="DELIVERYTYPE">DEPOSIT</mets:altRecordID>
<mets:altRecordID TYPE="DELIVERYSPECIFICATION">http://www.kb.se/namespace/digark/deliveryspecification/deposit/rssfeeds/rssfeeds.pdf</mets:altRecordID>
<mets:altRecordID TYPE="SUBMISSIONAGREEMENT">http://www.kb.se/namespace/digark/submissionagreement/31-KB669-2013</mets:altRecordID>
<mets:metsDocumentID>aip.mets.metadata</mets:metsDocumentID>
</mets:metsHdr>
- <mets:dmdSec ID="dmdSec001">
- <mets:mdWrap LABEL="Primary" MDTYPE="MODS">
- <mets:xmlData>
- <mods:mods>
  <mods:identifier type="uri">http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/</mods:identifier>
  <mods:identifier type="urn">urn:nbn:se:kb:dark-4457563</mods:identifier>
- <mods:name>
  <mods:namePart>Kim Malmgren</mods:namePart>
  - <mods:role>
    <mods:roleTerm type="code" valueURI="http://id.loc.gov/vocabulary/relators/cre" authority="marcrelator">cre</mods:roleTerm>
  </mods:role>
</mods:name>
- <mods:location>
  <mods:url usage="primary display">http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/</mods:url>
</mods:location>
- <mods:titleInfo>
  <mods:title>Personbil rammade busskur i Stockholm</mods:title>
</mods:titleInfo>
<mods:abstract><p>Polisen försökte stoppa bil som körde vårdslöst söder om Stockholm på torsdagskvällen.</p> <p>Färden slutade med att bilen rammade en busskur i
Farsta.</p> <p>En kvinna i 45-årsåldern befann sig i busskuren och fick flygas till sjukhus i ambulanshelikopter.</p> <p>Hennes tillstånd ska inte vara livshotande,
meddelar polisen under torsdagsnatten.</p></mods:abstract>
```

... and this ...

```
<mods:abstract><p>Polisen försökte stoppa bil som körde värdslöst söder om Stockholm på torsdagskvällen.</p> <p>Färden slutade med att bilen rammade en busskur i Farsta.</p> <p>En kvinna i 45-årsåldern befann sig i
busskuren och fick flygas till sjukhus i ambulanshelikopter.</p> <p>Hennes tillstånd ska inte vara livshotande, meddelar polisen under torsdagsnatten.</p></mods:abstract>
- <mods:originInfo>
  <mods:dateIssued encoding="w3cdtf">2015-02-13T00:46:52+01:00</mods:dateIssued>
  - <mods:place>
    <mods:placeTerm type="code" authority="marccountry">sw</mods:placeTerm>
  </mods:place>
</mods:originInfo>
<mods:accessCondition>gratis</mods:accessCondition>
<mods:typeOfResource>text</mods:typeOfResource>
<mods:genre authority="marcgt">article</mods:genre>
- <mods:language>
  <mods:languageTerm type="code" authority="iso639-2b">swe</mods:languageTerm>
</mods:language>
- <mods:physicalDescription>
  <mods:form authority="marcform">electronic</mods:form>
  <mods:digitalOrigin>born digital</mods:digitalOrigin>
</mods:physicalDescription>
- <mods:relatedItem type="host">
  <mods:identifier type="uri">http://libris.kb.se/resource/bib/11341783</mods:identifier>
</mods:relatedItem>
</mods:mods>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>
- <mets:dmdSec ID="dmdSec002">
  - <mets:mdWrap LABEL="Local" MDTYPE="MODS">
    - <mets:xmlData>
      - <mods:mods>
        - <mods:name type="corporate" valueURI="http://id.kb.se/organisations/SE5560254525" authority="local">
          - <mods:role>
            <mods:roleTerm type="code" valueURI="http://id.loc.gov/vocabulary/relators/pbl" authority="marcrelator">pbl</mods:roleTerm>
          </mods:role>
        </mods:name>
        - <mods:name type="corporate" valueURI="http://id.kb.se/organisations/SE5560254525" authority="local">
          - <mods:role>
            <mods:roleTerm type="text" authority="local">supplier</mods:roleTerm>
          </mods:role>
        </mods:name>
      </mods:mods>
    </mets:xmlData>
  </mets:mdWrap>
</mets:dmdSec>
- <mets:dmdSec ID="dmdSec003">
  <mets:mdRef xlink:href="http://libris.kb.se/resource/bib/17549561" LABEL="Libris" MDTYPE="MARC" LOCTYPE="URL"/>
</mets:dmdSec>
- <mets:amdSec ID="amdSec001">
  - <mets:techMD ID="techMD001">
    - <mets:mdWrap MDTYPE="PREMIS:OBJECT">
      - <mets:xmlData>
        - <premis:object xsi:type="premis:file" xsi:schemaLocation="info:lc/xmlns/premis-v2 http://www.loc.gov/standards/premis/v2/premis-v2-0.xsd">
          - <premis:objectIdentifier>
            <premis:objectIdentifierType>URN</premis:objectIdentifierType>
            <premis:objectIdentifierValue>urn:nbn:se:kb:dark-object-instance-25577832</premis:objectIdentifierValue>
          </premis:objectIdentifier>
        </premis:object>
      </mets:xmlData>
    </mets:techMD>
  </mets:amdSec>
```

... and more ...

```
<mods:abstract><p>Polisen försökte stoppa bil som körde värdslöst söder om Stockholm på torsdagskvällen.</p> <p>Färden slutade med att bilen rammade en busskur i Farsta.</p> <p>En kvinna i 45-årsåldern befann sig i
busskuren och fick flygas till sjukhus i ambulanshelikopter.</p> <p>Hennes tillstånd ska inte vara livshotande, meddelar polisen under torsdagsnatten.</p></mods:abstract>
- <mods:originInfo>
  <mods:dateIssued encoding="w3cdtf">2015-02-13T00:46:52+01:00</mods:dateIssued>
  - <mods:place>
    <mods:placeTerm type="code" authority="marccountry">sw</mods:placeTerm>
  </mods:place>
</mods:originInfo>
<mods:accessCondition>gratis</mods:accessCondition>
<mods:typeOfResource>text</mods:typeOfResource>
<mods:genre authority="marcgt">article</mods:genre>
- <mods:language>
  <mods:languageTerm type="code" authority="iso639-2b">swe</mods:languageTerm>
</mods:language>
- <mods:physicalDescription>
  <mods:form authority="marcform">electronic</mods:form>
  <mods:digitalOrigin>born digital</mods:digitalOrigin>
</mods:physicalDescription>
- <mods:relatedItem type="host">
  <mods:identifier type="uri">http://libris.kb.se/resource/bib/11341783</mods:identifier>
</mods:relatedItem>
</mods:mods>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>
- <mets:dmdSec ID="dmdSec002">
  - <mets:mdWrap LABEL="Local" MDTYPE="MODS">
    - <mets:xmlData>
      - <mods:mods>
        - <mods:name type="corporate" valueURI="http://id.kb.se/organisations/SE5560254525" authority="local">
          - <mods:role>
            <mods:roleTerm type="code" valueURI="http://id.loc.gov/vocabulary/relators/pbl" authority="marcrelator">pbl</mods:roleTerm>
          </mods:role>
        </mods:name>
        - <mods:name type="corporate" valueURI="http://id.kb.se/organisations/SE5560254525" authority="local">
          - <mods:role>
            <mods:roleTerm type="text" authority="local">supplier</mods:roleTerm>
          </mods:role>
        </mods:name>
      </mods:mods>
    </mets:xmlData>
  </mets:mdWrap>
</mets:dmdSec>
- <mets:dmdSec ID="dmdSec003">
  <mets:mdRef xlink:href="http://libris.kb.se/resource/bib/17549561" LABEL="Libris" MDTYPE="MARC" LOCTYPE="URL"/>
</mets:dmdSec>
- <mets:amdSec ID="amdSec001">
  - <mets:techMD ID="techMD001">
    - <mets:mdWrap MDTYPE="PREMIS:OBJECT">
      - <mets:xmlData>
        - <premis:object xsi:type="premis:file" xsi:schemaLocation="info:lc/xmlns/premis-v2 http://www.loc.gov/standards/premis/v2/premis-v2-0.xsd">
          - <premis:objectIdentifier>
            <premis:objectIdentifierType>URN</premis:objectIdentifierType>
            <premis:objectIdentifierValue>urn:nbn:se:kb:dark-object-instance-25577832</premis:objectIdentifierValue>
          </premis:objectIdentifier>
        </premis:object>
      </mets:xmlData>
    </mets:techMD>
  </mets:amdSec>
```

... and still more ... until:

```

    <premis:agentIdentifierType>MimerImplementation</premis:agentIdentifierType>
    <premis:agentIdentifierValue>se.kb.mimer.reception.processstepfactory.ValidateChecksums:release-3.1.1</premis:agentIdentifierValue>
  </premis:agentIdentifier>
  <premis:agentType>software</premis:agentType>
</premis:agent>
- <premis:agent>
  - <premis:agentIdentifier>
    <premis:agentIdentifierType>FileScannerImplementation</premis:agentIdentifierType>
    <premis:agentIdentifierValue>se.kb.mimer.droidscanner.FileScanner:release-1.0</premis:agentIdentifierValue>
  </premis:agentIdentifier>
  <premis:agentType>software</premis:agentType>
</premis:agent>
</mets:xmlData>
</mets:mdWrap>
</mets:digiprovMD>
</mets:amdSec>
- <mets:fileSec ID="fileSec001">
  - <mets:fileGrp ID="fileGrp001" USE="files">
    - <mets:file ID="file1" USE="files" SIZE="86825" OWNERID="urn:nbn:se:kb:dark-object-instance-25577832" MIMETYPE="text/html" CREATED="2015-02-13T11:29:05.472+01:00" CHECKSUMTYPE="MD5"
      CHECKSUM="007cd7d4dba36611d1e1c3a92f55ad80" ADMID="techMD001">
        <mets:Flocat xlink:type="simple" xlink:href="file:922c8f2254e061bc13af4b3216120311" LOCTYPE="URL"/>
        <mets:Flocat xlink:type="simple" xlink:href="file:http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/" LOCTYPE="URL" USE="originalURL"/>
      </mets:file>
    - <mets:file ID="file2" USE="files" SIZE="15475" OWNERID="urn:nbn:se:kb:dark-object-instance-25577833" MIMETYPE="image/jpeg" CREATED="2015-02-13T11:29:05.472+01:00" CHECKSUMTYPE="MD5"
      CHECKSUM="ca9e4bb972bba007ae403ad61b94451f" ADMID="techMD002">
        <mets:Flocat xlink:type="simple" xlink:href="file:68772a71f750a268d74711ee99c7153a.jpg" LOCTYPE="URL"/>
        <mets:Flocat xlink:type="simple" xlink:href="file:http://y.cdn-expressen.se/images/7a/cb/7acb8b54cabe4ff8bc4c04c4b9dbcb21/16x6/540@70.jpg" LOCTYPE="URL" USE="originalURL"/>
      </mets:file>
    </mets:fileGrp>
  - <mets:fileGrp ID="fileGrp002" USE="text/metadata">
    - <mets:file ID="file3" USE="text/metadata" SIZE="3919" OWNERID="urn:nbn:se:kb:dark-object-instance-25577834" MIMETYPE="text/xml" CREATED="2015-02-13T11:29:05.472+01:00" CHECKSUMTYPE="MD5"
      CHECKSUM="49f1e25cf1f8eb12bb1b0688a06a5370" ADMID="techMD003">
        <mets:Flocat xlink:type="simple" xlink:href="file:922c8f2254e061bc13af4b3216120311.rss.metadata" LOCTYPE="URL"/>
        <mets:Flocat xlink:type="simple" xlink:href="file:http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/.rss.metadata" LOCTYPE="URL" USE="originalURL"/>
      </mets:file>
    </mets:fileGrp>
  </mets:fileSec>
- <mets:structMap TYPE="physical" ID="structMap001">
  - <mets:div TYPE="files" ID="div001">
    - <mets:div TYPE="representation" ID="div002" ADMID="techMD003" DMDID="dmdSec001 dmdSec002">
      - <mets:div TYPE="maincontent" ID="div003">
        <mets:fptr FILEID="file1"/>
      </mets:div>
      - <mets:div TYPE="mediacontent" ID="div004">
        <mets:fptr FILEID="file2"/>
      </mets:div>
    </mets:div>
  - <mets:div TYPE="origmetadata" ID="div005">
    <mets:fptr FILEID="file3"/>
  </mets:div>
</mets:structMap>
</mets:mets>
```


So, how do we get from ... :

this ...

...to this

... to give access through:

LIBRIS 

[HELP](#) | [PÅ SVENSKA](#) | [PREFERENCES](#)

[Start](#) | [Extended search](#) | [Index A-Ö](#) | [Boolean](#) | [Subdatabases](#)

[My libraries](#)

Search: [onr:17549561](#) > Personbil rammade b...

1 of 1 ◀ Previous record | Next record ▶ [To hitlist](#)

[Overview](#) | [Details](#)

[LIBRIS Formathandbok](#) (Information om MARC21) [Details](#) | [MARC](#)

000 01137cam a22002773a 4500
001 17549561
003 SE-LIBR
005 20150330152918.0
007 cr|||||||||
008 150213s2015 sw |||| o |||| ||swe c
024 7 a http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/2 uri
040 a S
041 0 a swe
042 9 EPLK
245 1 0 a Personbil rammade busskur i Stockholm/h [Elektronisk resurs]
260 c 2015-02-12T21:09:31+01:00
506 0 a gratis
520 a Polisen försökte stoppa en bil som körde vårdslöst söder om Stockholm. Färden slutade med att bilen rammade en busskur i Farsta. En person som befann sig i busskuren har flugits till Karolinska universitetssjukhuset i ambulanshelikopter.
655 7 a article2 marcgt
772 1 8 i channel recordw 14749891
773 0 w 11341783
856 4 0 u http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/
841 5 APISa x ab 150213||0000|||||001|||||000000e 1
024 7 5 APISa urn:nbn:se:kb:dark-4457563.2 urn
852 5 APISb APIS
856 4 0 5 APISu http://urn.kb.se/resolve?urn=urn:nbn:se:kb:dark-4457563

... or – presently - rather this:

MARC | System | History

Leader 00925cam a22002293a 4500 006

005: 20150213132541.0 007 c:rl|||||||

008 150213 s 2015 __ sw_ ||| _ o _ |||| _ || swe _ c

Tag	I1	I2	Subfield Data
→ 024	7		⌘a http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/ ⌘2 uri
040			⌘a S
041	0		⌘a swe
042			⌘9 EPLK
245	1	0	⌘a Personbil rammade busskur i Stockholm ⌘h [Elektronisk resurs]
260			⌘c 2015-02-12T21:09:31+01:00
506	0		⌘a gratis
520			⌘a Polisen försökte stoppa en bil som körde vårdslöst söder om Stockholm. Färden slutade med att bilen rammade en busskur i Farsta. En person som befann sig i busskuren har flugits till Karolinska universitetssjukhuset i ambulanshelikopter.
599			⌘a SUPPRESSRECORD
655		7	⌘a article ⌘2 marcgt
772	1	8	⌘i channel record ⌘w 14749891
773	0		⌘w 11341783
856	4	0	⌘u http://www.expressen.se/nyheter/personbil-rammade-busskur-i-stockholm/

No information found

13:26

Stumbling blocks:

1. Negative expectations
2. The Law - what can we legally ask for?
3. Positive expectations – too much, too soon!
4. The Publishers
5. Format / metadata standard constraints
6. Deduplication and version control
7. Inundation



Stumbling blocks:



1. Negative expectations

- Swedish Media Publishers: “new law on e-legal deposit too complicated and resource demanding”
- Legislator argument: harvesting Swedish web pages not enough (already in effect since 1997), no access to content protected by passwords, encrypted etc.
- Result: lower requirements for metadata on e-legal deposit

Stumbling blocks:



2. The Law – what can we legally ask for?

- Default method of e-legal deposit delivery: physical carrier (USB-stick)
- NLS accorded the right to decide about other possible methods of delivery of e-legal deposit, viz. electronic file transfer
- Presently four different methods of online delivery: web form, FTP, OAI-PMH and RSS-feeds
- Self-interest: avoid too strong metadata requirements on suppliers choosing online delivery, in order not to make them prefer the USB-stick instead
- Strike a balance between requirements of the law, the metadata standard itself, a sufficiently rich metadata quality and a fair enough work load on the supplier

Stumbling blocks:



3. Positive expectations: too much, too soon

- E-legal deposit replacing earlier resource demanding in-house manual library cataloging with automated, machine generated catalog records
- Mimer already produces machine generated high quality catalog records in LIBRIS, comparable to corresponding manually cataloged records.
- Requires substantial initial effort from the supplier, setting up and managing their system for delivery, including metadata.
- End users: expecting free, ubiquitous access to copyrighted documents, otherwise restricted by payment or password control? Disillusionment!
- Substantial legal, financial and technical issues to deal with before that can happen.
- NLS only just begun to deal with technical issues (DIPs, GUI, search options etc.)

Stumbling blocks:



4. The Publishers

- News media publishers are active in a very volatile and vulnerable market -> accelerating rate of mergers, ownership changes and discontinuances -> keeping track of online news media publishers a real challenge.
- We need publisher identifiers that are persistent, unique and distinct from the supplier-IDs, which are often shared by a number of online news media belonging to the same publishing house or media syndicate.

Stumbling blocks:



5. Format / metadata standard constraints

- RSS 2.0 designed to be a very simple and easy to use standard, with very few mandatory elements or attributes
- NLS and Mimer uses possibility to “mix and match” by adding certain elements from MediaRSS and Dublin Core (*dcterms*) as mandatory in our implementation of RSS for e-legal deposit
- 7 mandatory elements and 3 mandatory if applicable.
- Most mandatory elements derived from law requirements: identifier (*guid*), internet address of resource and constituent files (*link*, *media:content/@url*), publishing date (*pubDate*), publisher (*dcterms:publisher*), accessibility at the time of publishing (*dcterms:accessRights*), file format of resource and constituent files (*dcterms:format*, *media:content/@type*)
- Mandatory *title* element - a byproduct of the RSS format itself: “... at least one of title or description must be present.” -> made *title* mandatory, avoid never knowing beforehand which one we would get.
- Among optional metadata elements: *license*, statement of responsibility (*creator*, *contributor*), *keywords*, *categories* (subject headings) - considered to be less important for news feeds, a price to pay for keeping it simple

Stumbling blocks:



6. Deduplication and version control

- Format limitation of RSS 2.0 specification: main identifier for items, *guid*, can contain almost anything, since data type is not specified -> makes it very difficult to use for deduplication and version control searches
- Recently trying to overcome this limitation by means of optional and repeatable *dcterms:identifier* element, with an *xsi:type* attribute for specification of identifier type (such as, e.g. *doi*, *hdl*, etc.), allowing effective deduplication
- Still dependent on the supplier / publisher to use *the same* identifier value, the same *guid*, for every update of one and the same item (misfortune cases with corrected spelling in update)

The last hurdle?



7. Inundation

- Forces of nature: too many online news items out there. Even with complete deduplication and version control, still risk of completely inundating library catalog LIBRIS with records for news items, other document records being submerged in a sea of news bites.

- “Solution” (temporary): suppression of news item records (‘articles’) from display in LIBRIS web catalog.

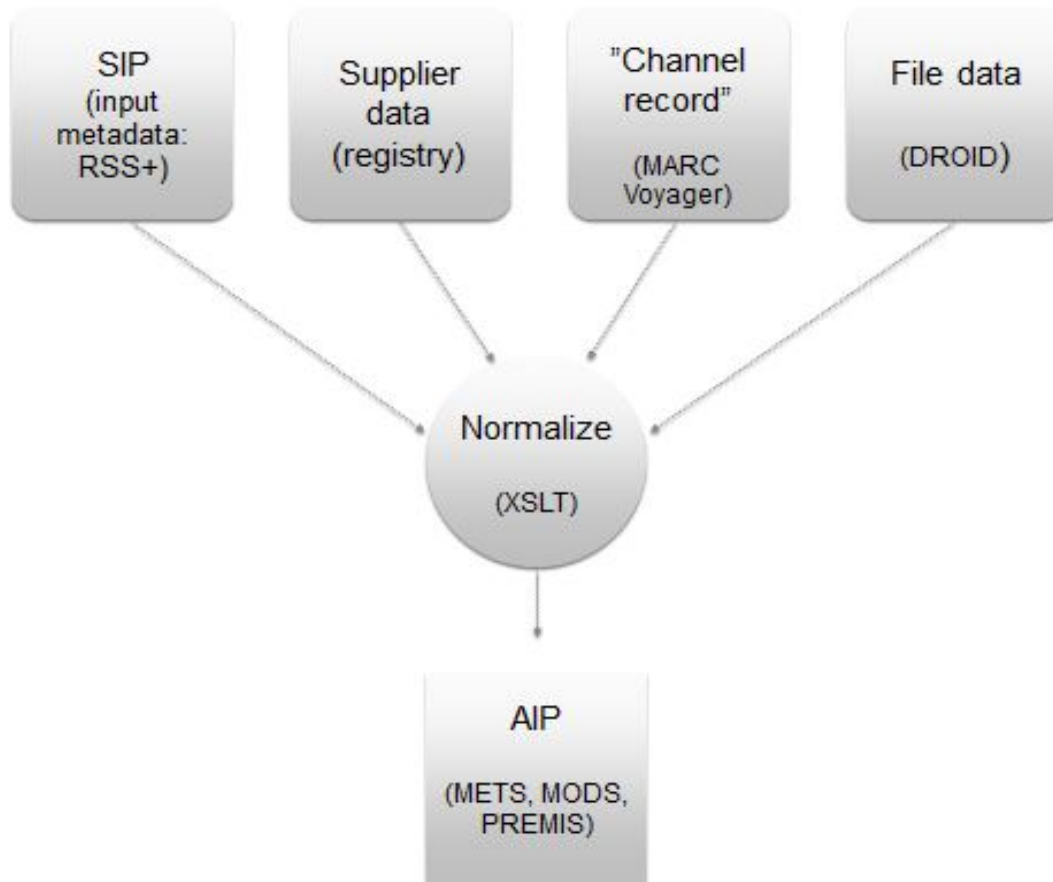


Processing

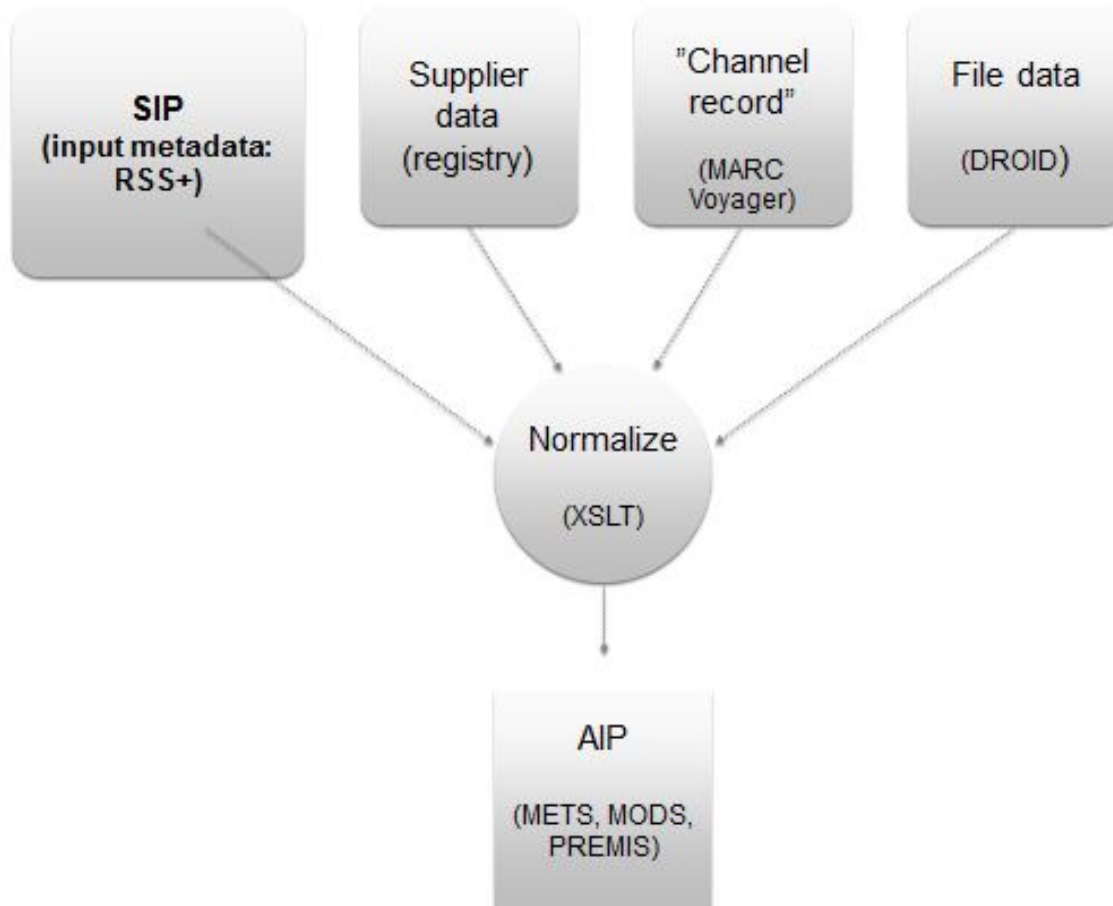
RSS:

- 1. Feed-reader & fetching the files
- 2. Validation
- 3. Split up of items
- 4. (Re)packaging into SIPs (input metadata + files)
- 5. Validation
- 6. **Normalization:** transformation to AIP & enrichment (XSLT)
- 7. Transformation of AIP to LIBRIS catalog (MARCXML).

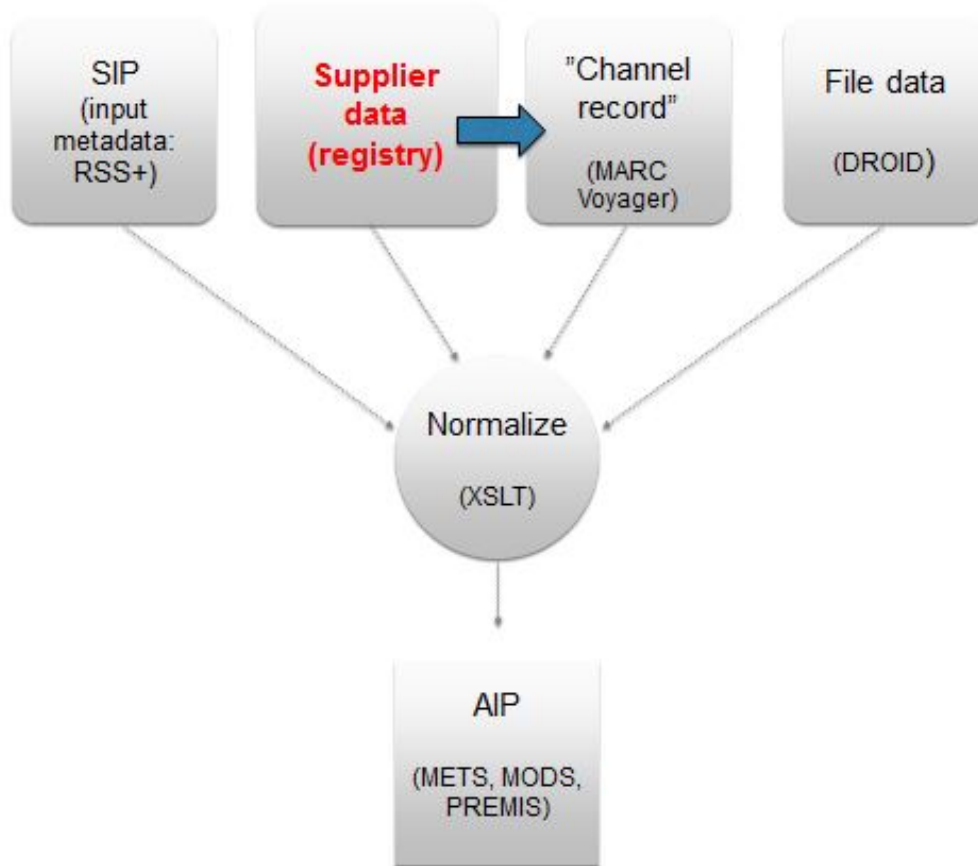
Normalization



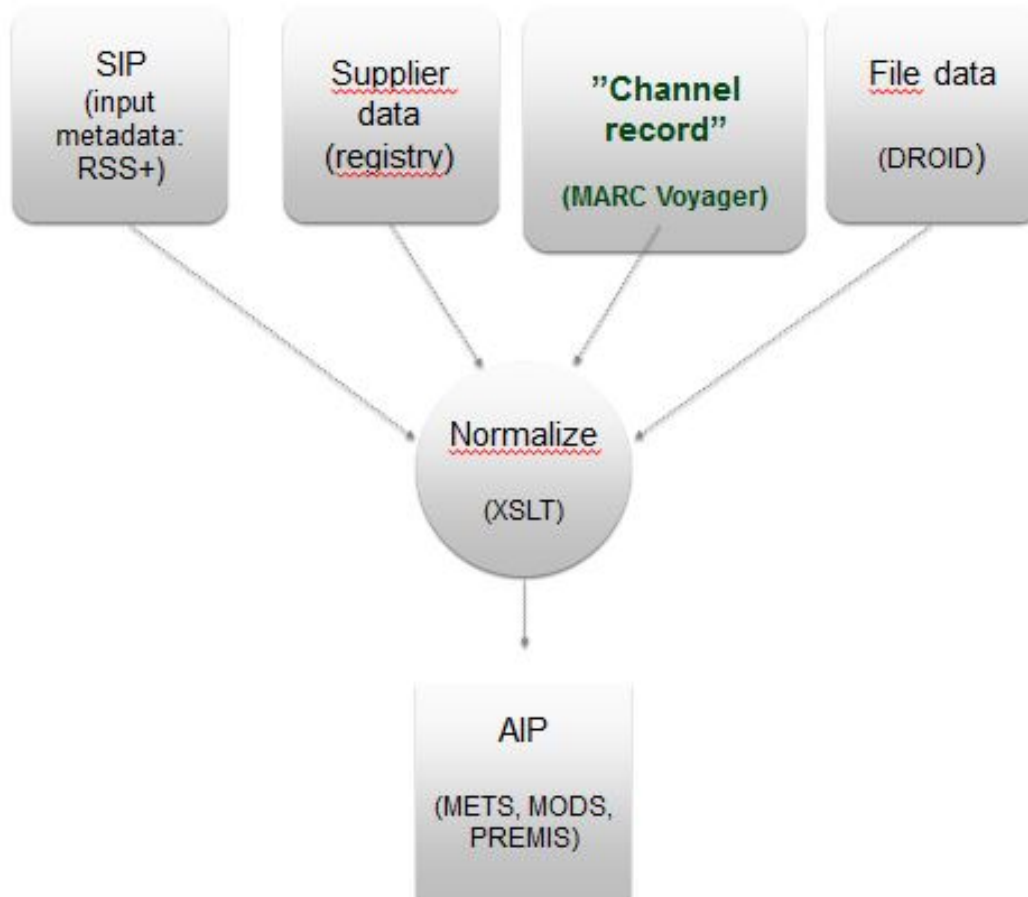
Normalization 1



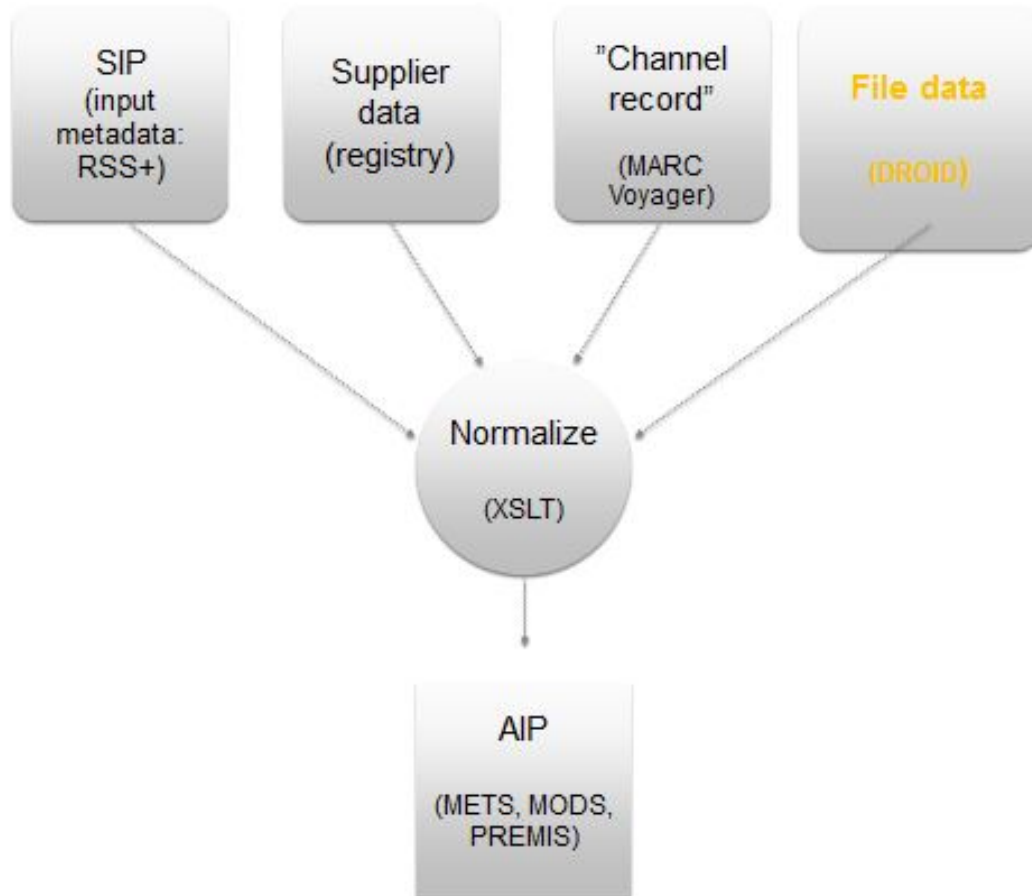
Normalization 2



Normalization 3



Normalization 4



Preservation

- No legal right to specify preferred file formats. Mimer must accept and ingest all types of electronic resources in file formats published online.
- For preservation planning and to apply the right measures at the right time (e.g. migration), need to know which data formats are in store.
- File data: as a minimum: MIME types.
- File validation using DROID: format name, format version and key in format register PRONOM.
- Format identifiers and checksums (if any) from supplier also saved.
- Mimer always adds its own checksums (MD5).
- Preservation metadata standard: PREMIS; information about actions (events) performed on each data file (validation, normalization), results of actions, applications software (agents) that performed the events

Giving access

Only preliminary and preparatory steps:

- Transformation of AIP metadata to MARC catalog records in LIBRIS, the national library union catalog, by means of XSLT.
- Preparation for LIBRIS XL, new cataloging client and web catalog based entirely on linked data (JSON-LD) and discarding the MARC format, will eventually require new transformation of e-legal deposit (MODS-RDF?)
- Hoping for new LIBRIS XL web catalog to allow display also of now suppressed catalog records of online news items (web articles)
- Preparatory development of possible DIPs
- Query processing and access times have been speeded up considerably through transfer from Fedora triple store to XML database solution (MarkLogic)

The future?

- Offering other methods of delivery and other metadata formats for e-legal deposit (e.g. Atom, RDFa, ONIX?).
- Resolution of legal and financial issues (copyright, licensing)?
- Being prepared for the unexpected

Time for a break?



not quite **THE END ... but ,**

Thank You! for now