

# Web servisai

## SOAP serializavimas

Osvaldas Grigas

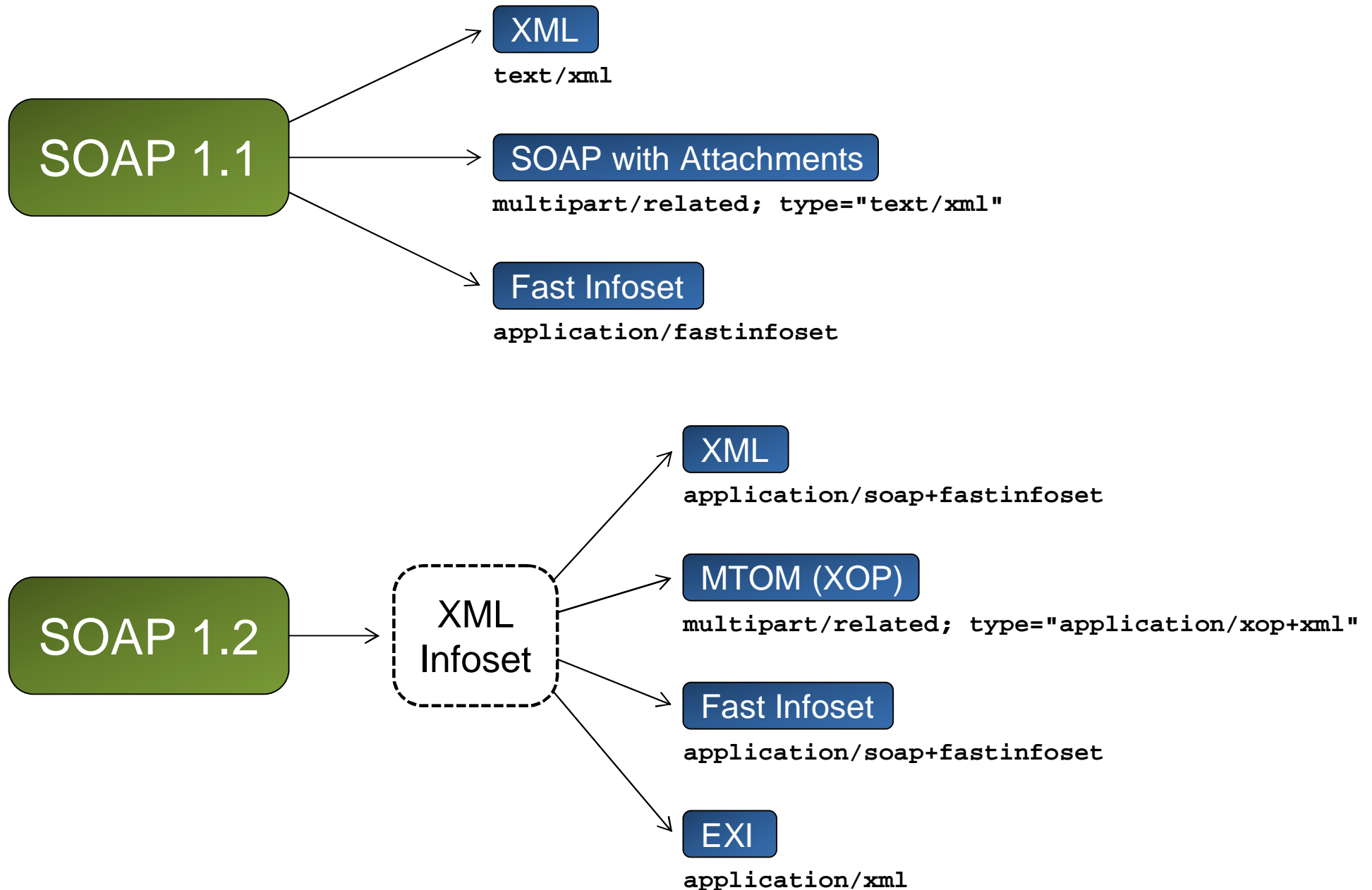
# Serializavimas

- *Serializavimas* - tai duomenų struktūros transformavimas į baitų seką, kurią galima saugoti diske arba persiųsti tinklu.
- *Deserializavimas* - atvirkštinis procesas, t.y. kai iš baitų sekos atstatoma originali duomenų struktūra.
- SOAP pranešimas paprastai konstruojamas programos atmintyje (naudojant DOM ar kokį kitą XML API), tada serializuojamas į tekstinį XML formatą ir siunčiamas tinklu. Tai tinka daugeliui atvejų, bet ne visada.
  - Tekstinio XML apdorojimas užtrunka.
  - Stambūs pranešimai, koduoti tekstiniu XML, lėtai keliauja tinklu.
  - Tekstinis XML nepritaikytas dvejetainiams duomenims siųsti. Norint tokiu būdu nusiųsti dvejetainį failą, tenka jį koduoti tekstiniais simboliais (*base64*) - dėl to jis padidėja 33%.

# Dvejetainiai failai per SOAP

- Tarkim, SOAP pranešime mums reikia siųsti dvejetainį failą. Galimi sprendimai:
  - A. Užkoduoti failo turinį *base64* koduote ir įdėti į XML elementą (neefektyvu).
  - B. Iškelti failą iš SOAP pranešimo ir serializuoti atskirai.
  - C. Serializuoti visą SOAP pranešimą dvejetainiu formatu.
- *SOAP 1.2* remiasi XML Infoset specifikacija, todėl jį galima natūraliai serializuoti bet koku dvejetainiu XML formatu (XOP, Fast Infoset, EXI ir kt.).
- Su *SOAP 1.1* sunkiau, nes jis remiasi tekstiniu XML. Yra sukurti specialūs pranešimų formatai, skirti efektyviam dvejetainio turinio siuntimui su SOAP 1.1.

# SOAP serializavimas





XML

# XML serializavimas

- MIME tipas
  - SOAP 1.1 pranešimuose: *text/xml* (kartais *application/xml*)
  - SOAP 1.2 pranešimuose: *application/soap+xml*
- Dažniausiai naudojamas ir plačiausiai palaikomas SOAP serializavimo būdas
- Naudoja tekstinį XML 1.0 formatą
- Labiausiai tinka smulkiems pranešimams
- Dvejetainis turinys paprastai koduojamas *base64*

# Dvejetainis turinys XML dokumente

HTTP/1.1 200 OK

Content-Type: *application/soap+xml*

```
<?xml version="1.0"?>
```

```
<soap:Envelope xmlns:soap="http://www.w3.org/2003/05/soap-envelope">
```

```
  <soap:Body>
```

```
    <image format="JPEG" xmlns="urn:image-service">
```

```
      <title>Mano senelio nuotrauka</title>
```

```
      <!-- dvejetainiai duomenys koduojami base64 -->
```

```
      <data>
```

```
        FylHBhc3Npb24gZnJvbSBvdGhlciBhbmltYWxzLCB3aGljaCBpcy  
        QsIHRoYXQgYnkgYSBwZXJzZXZlcmFuY2Ugb2YgZGVsaWdod  
        BpbmRIZmF0aWdhYmxlIGdlbmVvYXRpb24gb2Yga25vd2xIZGd  
        B2ZWhlbWVvY2Ugb2YgYW55IGNhcm5hbCBwbGVhc3VyZS4=
```

```
      </data>
```

```
    </image>
```

```
  </soap:Body>
```

```
</soap:Envelope>
```

# SOAP with Attachments



# SOAP with Attachments

- Kartais trumpinamas "SwA"
- Naudojamas SOAP 1.1 pranešimams, kuriuose reikia siųsti dvejetainius failus
- Atskiria dvejetainius fragmentus nuo SOAP XML
  - SOAP dalį koduoja tekstiniu XML
  - Dvejetainius fragmentus koduoja tiesiog baitais
- Dvejetainiai fragmentai vadinami priedais (*attachments*)
  - Bet galima naudoti ir tekstinius priedus
- Šį formatą Java kalboje palaiko SAAJ (SOAP with Attachments API for Java)
- SOAP 1.2 pranešimuose SwA nenaudojamas, nes neatitinka XML Infoset specifikacijos

# SwA = SOAP + MIME

- MIME yra standartinis būdas tinklu keliaujančiam turiniui formatuoti ir jo tipui aprašyti
  - nors pradžioje MIME buvo naudojamas tik el. laiškas koduoti
- Kiekvienam turinio tipui suteikiamas unikalus pavadinimas (*MIME type*), kurį sudaro *tipas* ir *potipis*, atskirti simboliu "/"
  - pvz. tekstinis turinis - *text/plain*, JPEG paveikslėlis - *image/jpeg* ir t.t.
- MIME taip pat palaiko *daugialypį* turinį, kurį sudaro keli skirtingų tipų komponentai. Pvz. laiškas su priedais yra daugialypis pranešimas, kurį gali sudaryti:
  - formatuotas tekstas (Content-Type: text/html)
  - paveikslėlis (Content-Type: image/gif)
  - MS Word priedas (Content-Type: application/msword)
  - ir kt.

# Multipart MIME

- Daugialypio pranešimo pradžioje aprašomas skirtukas (*boundary*), kuris toliau naudojamas pranešimo dalims atskirti, pranešimo pradžiai ir pabaigai žymėti. Skirtukas paprastai būna ilgas ir sudėtingas.
- Kiekvienos dalies turinio tipą aprašo jos header'is *Content-Type*

MIME-version: 1.0

Content-Type: *multipart/mixed*; boundary="**skirtukas**"

**--skirtukas**

Content-Type: *text/plain*

Sveikas, siunčiu tau dokumentą, kurio prašei.  
Iki greito!

**--skirtukas**

Content-Type: *application/msword*

Content-Transfer-Encoding: binary

jÃ0þï...½fÑ}QÒÃ□%v/¥□C/£} á(□h"□û□ëÙOÇ□»□„,¤ï÷©=þ®‹ùá"ç ¢šª  
µà□□â£x}rĭ□x□%o%o¢□œwçØ□îrĬ:\TZaGó\*™y8lÂjæbRÆíc|XÁ»‹□Ç¿½  
ÎÉÅ‹4vfaÇÖvl;©Á³'S□†ÆùAÆ8Æ|Ò\*□uâ£{àè-,ßÿ0%M0Á7%□jõ□~<€ä  
^ÙO¼R®BdÑÉ□ÒJEÛ4b\$□§'q\_x□ž□à6LÓõ¤R×7`®□"Éÿ³Ã0ìžOÁ□\_ v  
**--skirtukas--**

# Multipart MIME

Daugialypis turinys pats turi MIME tipą. Galimi netgi keli tipai:

- *multipart/mixed*
  - Jį sudaro keli komponentai, kurių turinys nesusijęs.
  - Pvz. laiško tekstas, "MS Word" priedas ir "MS Excel" priedas.
- *multipart/alternative*
  - Jį sudaro to paties turinio kelios versijos arba keli formatai.
  - Pvz. laiškas paprastu tekstu ir tas pats laiškas HTML formatu.
- *multipart/related*
  - Jį sudaro vienas "šakninis" komponentas ir keli jo "vaikai", kurie savo ruožtu yra "šakninio" komponento dalys.
  - Kiekvienas komponentas yra unikaliam identifikuojamas, todėl "šakninis" komponentas gali savo viduje laikyti nuorodas į "vaikus".
  - Pvz. laiškas HTML formatu su paveikslėliais. Paveikslėlio komponentas galėtų turėti header'į *Content-ID: 12345*, o HTML komponentas laikytų nuorodą į paveikslėlį per tokią ar panašią žymę: ``
- *ir kiti multipart potipiai...*

# SOAP with Attachments

- SOAP with Attachments yra SOAP pranešimas, siunčiamas MIME formatu.
- Tai *multipart/related* tipo pranešimas, kurio "šakninis" komponentas visada yra SOAP Envelope (*Content-Type: text/xml*).
- Bet kurio XML elemento turinį galima išimti iš SOAP dalies ir iškelti į atskirą MIME komponentą.
- Kiekvienam komponentui suteikiamas unikalus *Content-ID*
- Taip iškėlus fragmentą, SOAP dalyje lieka tuščias XML elementas. Jam suteikiamas atributas *href* su reikšme "cid:xxx", kur xxx - iškelto turinio *Content-ID*.
- Tokiu būdu galima atskirti dvejetainius duomenis nuo SOAP pranešimo ir juos efektyviai užkoduoti baitais.
- *href* nuorodos leidžia gavėjui rekonstruoti originalią pranešimo struktūrą.

# SOAP with Attachments pavyzdys

HTTP/1.1 200 OK

Content-Type: *multipart/related*; type="text/xml"; boundary="skirtukas"; start="<SOAP\_PART>"

**--skirtukas**

Content-Type: *text/xml*

Content-ID: <SOAP\_PART>

<?xml version="1.0"?>

<soap:Envelope xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">

<soap:Body>

<image format="JPEG" xmlns="urn:image-service">

<title>Mano senelio nuotrauka</title>

<data href="cid:IMAGE\_PART"/>

</image>

</soap:Body>

</soap:Envelope>

**--skirtukas**

Content-Type: *image/jpeg*

Content-Transfer-Encoding: binary

Content-ID: <IMAGE\_PART>

µà□□å£x}rī□x□%o%oç□œwçØ□îrî:\TZaGó\*™y8IÂjæbRÆíc|XÅ»◁□Ç¿½  
ÎÉÅ◁4vfaÇÖvlj©Á³'S□†ÆùAÆ8Æ|Ò\*□uâ£{àè-,ßÿ0%M0Á7%□jõ□~<€ä  
^ÙO¼R®BdÑÉ□ÒJEÛ4b\$□§'q\_x□~ž□à6LÓõ¤R×7`®□"Éÿ³Ã0ìžOÁ□~v

**--skirtukas--**



MTOM

# MTOM ir XOP

- XOP = XML-binary Optimized Packaging
  - Vienas iš XML Infoset serializavimo formatų
  - Tinka bet kokiam XML dokumentui (ne tik SOAP)
  - XOP = XML + MIME
  - Naudoja *multipart/related* MIME formatą
  - Leižia nuo XML dokumento atskirti dvejetainius fragmentus, iškelti į atskiras MIME pranešimo dalis ir užkoduoti baitais
- MTOM = Message Transmission Optimization Mechanism
  - MTOM = SOAP + XOP
  - MTOM yra XOP pritaikymas SOAP pranešimams serializuoti
  - MTOM specifikacija yra labai glausta, nes visos serializavimo/deserializavimo detalės pateikiamos XOP specifikacijoje
  - XOP ir MTOM yra W3C rekomendacijos
  - Java kalboje MTOM mechanizmą palaiko JAX-WS 2.0



# MTOM vs. SwA

- MTOM išoriškai panašus į SOAP with Attachments, tačiau...
- MTOM būdu serializuotas SOAP pranešimas yra XML dokumentas, o SwA - nėra. MTOM naudoja XOP, kuris leidžia serializuoti bet kokį XML Infoset'ą (ne vien tik SOAP).
- SwA tinka SOAP 1.1, o MTOM - SOAP 1.2 pranešimams
- Ryšiams tarp MIME dalių išlaikyti SwA naudoja atributą *href*, kuris gali konfliktuoti su jau egzistuojančiu "href" atributu SOAP pranešime. MTOM tam tikslui naudoja elementą *Include* iš XOP vardų erdvės, pavyzdžiui:

```
<data xmlns:xop="http://www.w3.org/2004/08/xop/include">  
  <xop:Include href="cid:12345" />  
</data>
```

# MTOM pavyzdys

HTTP/1.1 200 OK

Content-Type: *multipart/related*; type="application/xop+xml"; boundary="skirtukas"; start="<SOAP\_PART>"

**--skirtukas**

Content-Type: *application/xop+xml*; type="application/soap+xml"

Content-ID: <SOAP\_PART>

<?xml version="1.0"?>

<soap:Envelope xmlns:soap="http://www.w3.org/2003/05/soap-envelope">

<soap:Body xmlns:xop="http://www.w3.org/2004/08/xop/include">

<image format="JPEG" xmlns="urn:image-service">

<title>Mano senelio nuotrauka</title>

<data><xop:Include href="cid:IMAGE\_PART"/></data>

</image>

</soap:Body>

</soap:Envelope>

**--skirtukas**

Content-Type: *image/jpeg*

Content-Transfer-Encoding: binary

Content-ID: <IMAGE\_PART>

µà□□å£x}rĭ□x□%o%oç□œwçØ□îrĭ:\TZaGó\*™y8IÂjæbRÆíc|XÅ»◁□Ç¿½  
ÎÉÅ◁4vfaÇÖvlĭ©Á³S□†ÆùAÆ8Æ|Ò\*□uâ£{àè-ßŸ0%M0Á7%□jõ□~<€ä  
^ÛO¼R®BdÑÉ□ÒJEÚ4b\$□\$‘q\_x□~ž□à6LÓõ¤Rx7`®□“Éÿ³Ã0ÏžOÁ□\_v

**--skirtukas--**



Kiti formatai

# Fast Infoset

- Fast Infoset - visiškai dvejetainis XML Infoset serializavimo formatas
- ISO standartas
- Java kalboje jį palaiko JAX-RPC ir JAX-WS bibliotekos
- Gali būti naudojamas bet kokiai SOAP versijai
- Ne tik sumažina pranešimo dydį, bet ir žymiai pagreitina jo apdorojimą, todėl kartais naudojamas ir tekstiniams pranešimams optimizuoti
- MIME tipas:
  - SOAP 1.1 pranešimams: *application/fastinfoset*
  - SOAP 1.2 pranešimams: *application/soap+fastinfoset*

# EXI

- Efficient XML Interchange (EXI) - eksperimentinis visiškai dvejetainis XML Infoset serializavimo formatas
- Šiuo metu tik W3C projektas, bet turėtų tapti rekomendacija
- Kol kas nestandardizuotas ir mažai naudojamas
- W3C atliko išsamų įvairių dvejetainių XML formatų tyrimą, ir nusprendė, kad AgileDelta Efficient XML formatas yra efektyviausias
- EXI yra kompaktiškesnis nei Fast Infoset, nes sugeba optimizuoti kompresiją atsižvelgiant į dokumento XML schema (nors gerai veikia ir be schemos)
- Gali būti naudojamas SOAP 1.2 pranešimams optimizuoti
- MIME tipas: *application/xml*, nes laikoma, kad EXI dokumentas yra tiesiog XML dokumentas su EXI kodavimu