

# A born digital irodalmi hagyatékok és az e-mailek archiválása a közgyűjteményekben

Szűcs Kata Ágnes

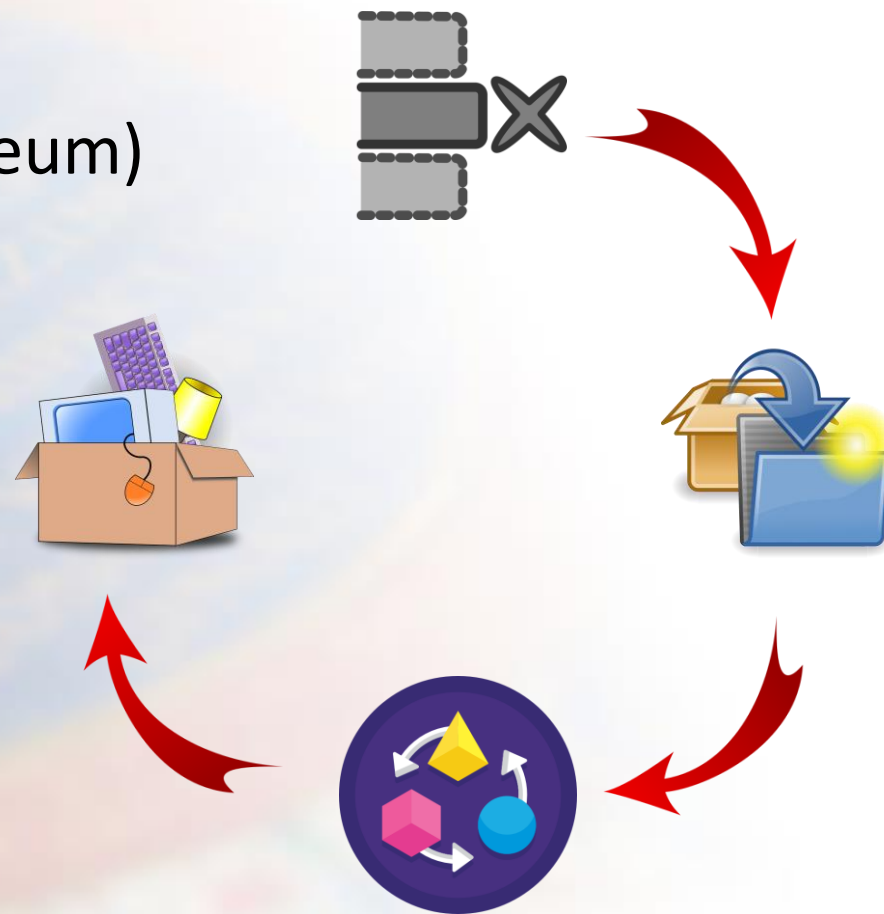
Szakkönyvtári Seregszemle

2022. március 29.



# Előzmények

- Hangoskönyv projekt (Petőfi Irodalmi Múzeum)
- Intézményi sajátosságok
  - Informatikai rendszerek
  - Microsoft Office környezet (PST)
- Workflow lépései
  1. válogatás és rendszerezés
  2. exportálás
  3. konverzió
  4. csomagolás → Baglt





# Kapcsolódó szabványok



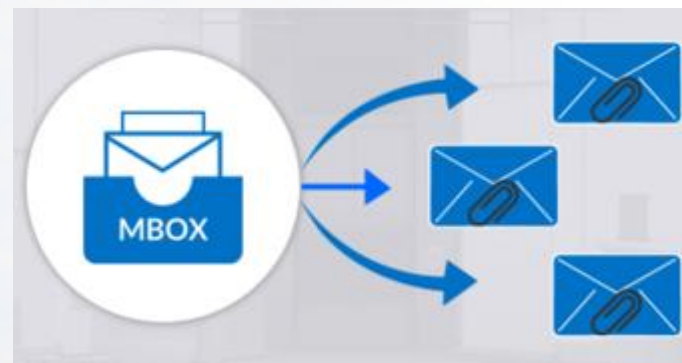
- **Internet Engineering Task Force (IETF)**
- **Request for Comments (RFC).**
- Az üzenetformátum
  - RFC 5322 (IMF – Internet Message Format),
- Az üzenetküldés
  - RFC 5321 (SMTP – Simple Mail Transfer Protocol)
- A hozzáférést szabályozó protokollok
  - Internet Message Access Protocol (IMAP)
  - Post Office Protocol v3 (POP3)
  - Multipurpose Internet Mail Extensions (MIME).

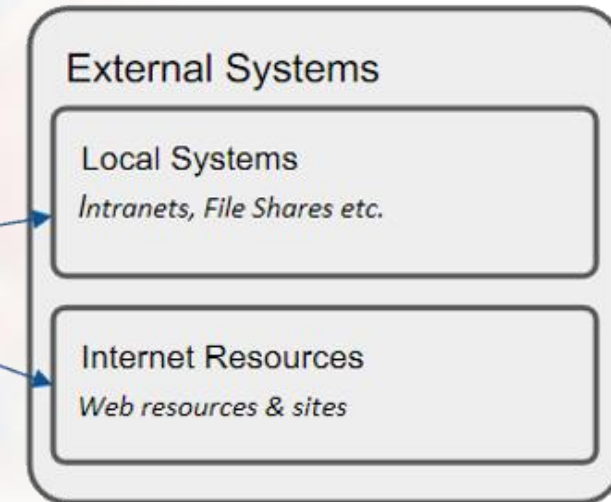
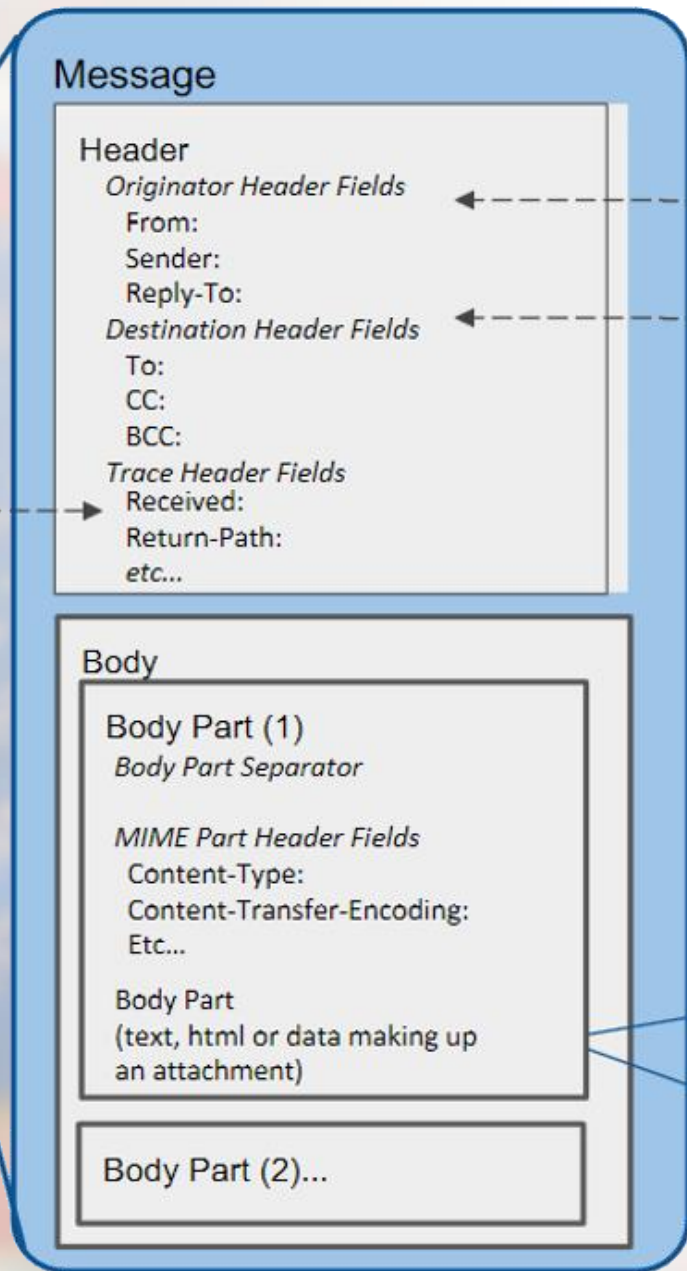
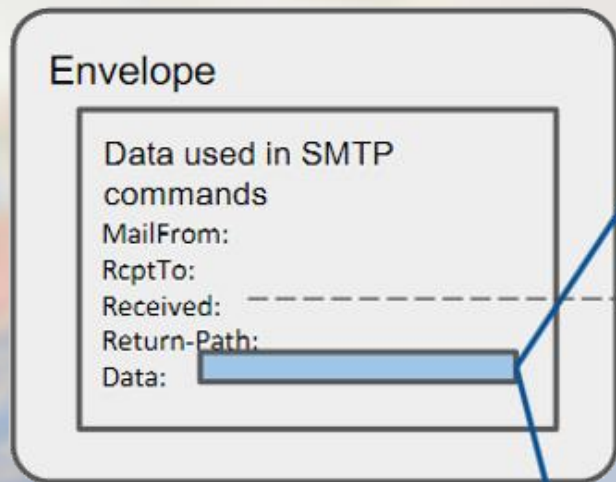
# Szabványos formátumok

.eml

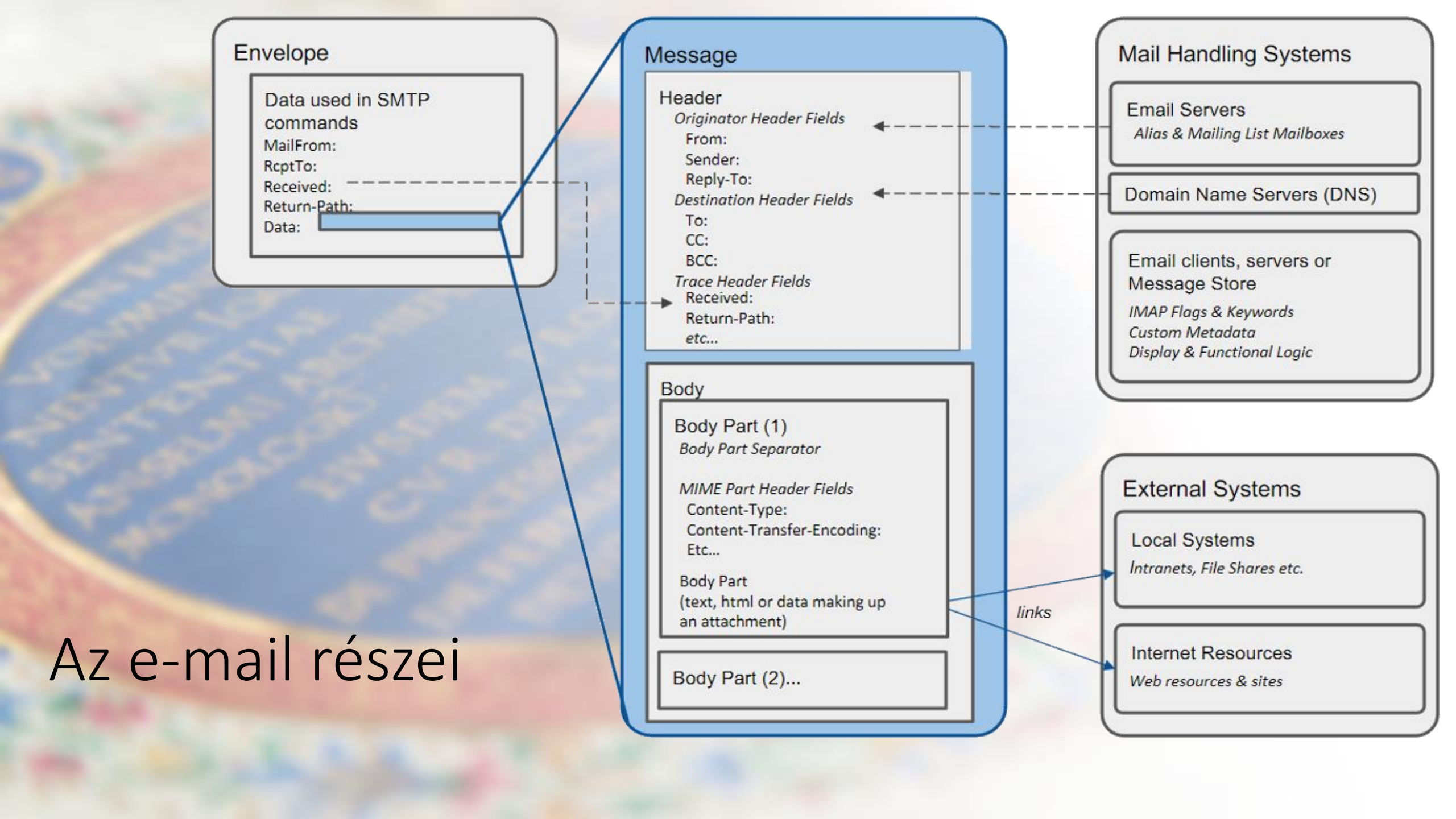


.mbox





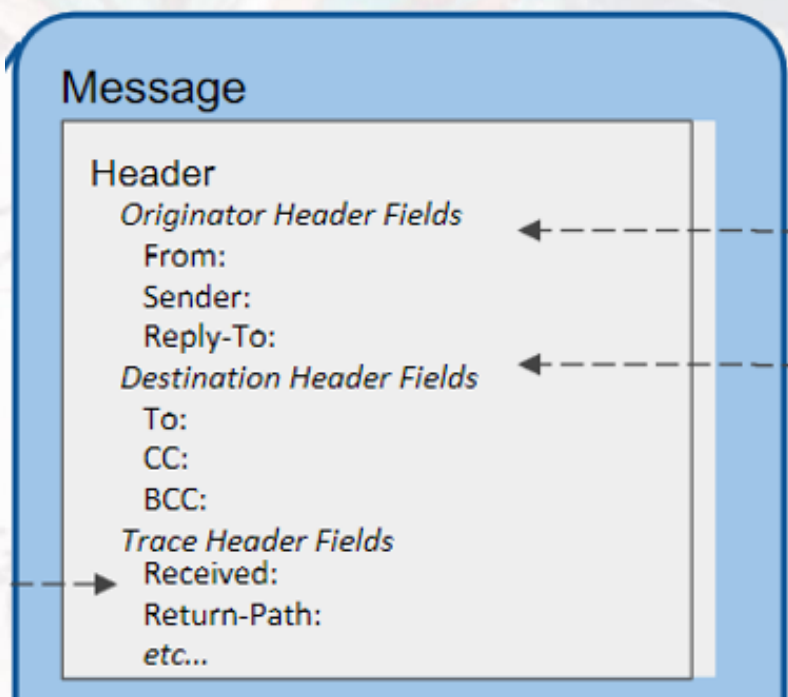
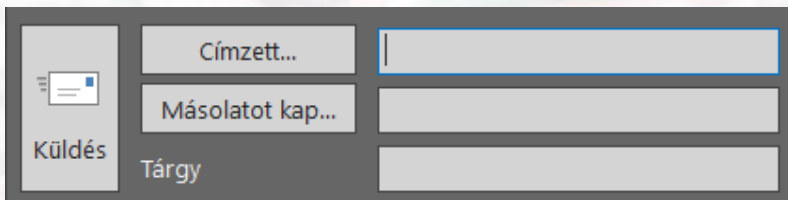
Az e-mail részei





</header>

# RFC5322 szerinti header fiek



## dátumra vonatkozó:

- orig-date = "Date:" date-time

## feladóra vonatkozó:

- from = "From:" mailbox-list
- sender = "Sender:" mailbox
- reply-to = "Reply-To:" address-list

## címzetre vonatkozó:

- to = "To:" address-list
- cc = "Cc:" address-list
- bcc = "Bcc:" [address-list / CFWS]

## identifikációs:

- message-id = "Message-ID:" msg-id
- in-reply-to = "In-Reply-To:" 1\*msg-id
- references = "References:" 1\*msg-id
- msg-id = [CFWS] "<" id-left "@" id-right ">"  
[CFWS]
- rid-left = dot-atom-text / obs-id-left

- id-right = dot-atom-text / no-fold-literal / obs-id-right

- no-fold-literal = "[" \*dtext "]"

## információs:

- subject = "Subject:" unstructured
- comments = "Comments:" unstructured
- keywords = "Keywords:" phrase \*(", " phrase)

## továbbított üzenet header mezői:

- resent-date = "Resent-Date:" date-time
- resent-from = "Resent-From:" mailbox-list
- resent-sender = "Resent-Sender:" mailbox
- resent-to = "Resent-To:" address-list
- resent-cc = "Resent-Cc:" address-list
- resent-bcc = "Resent-Bcc:" [address-list / CFWS]
- resent-msg-id = "Resent-Message-ID:" msg-id

trace mezők: [...]

opcionális mezők: [...]



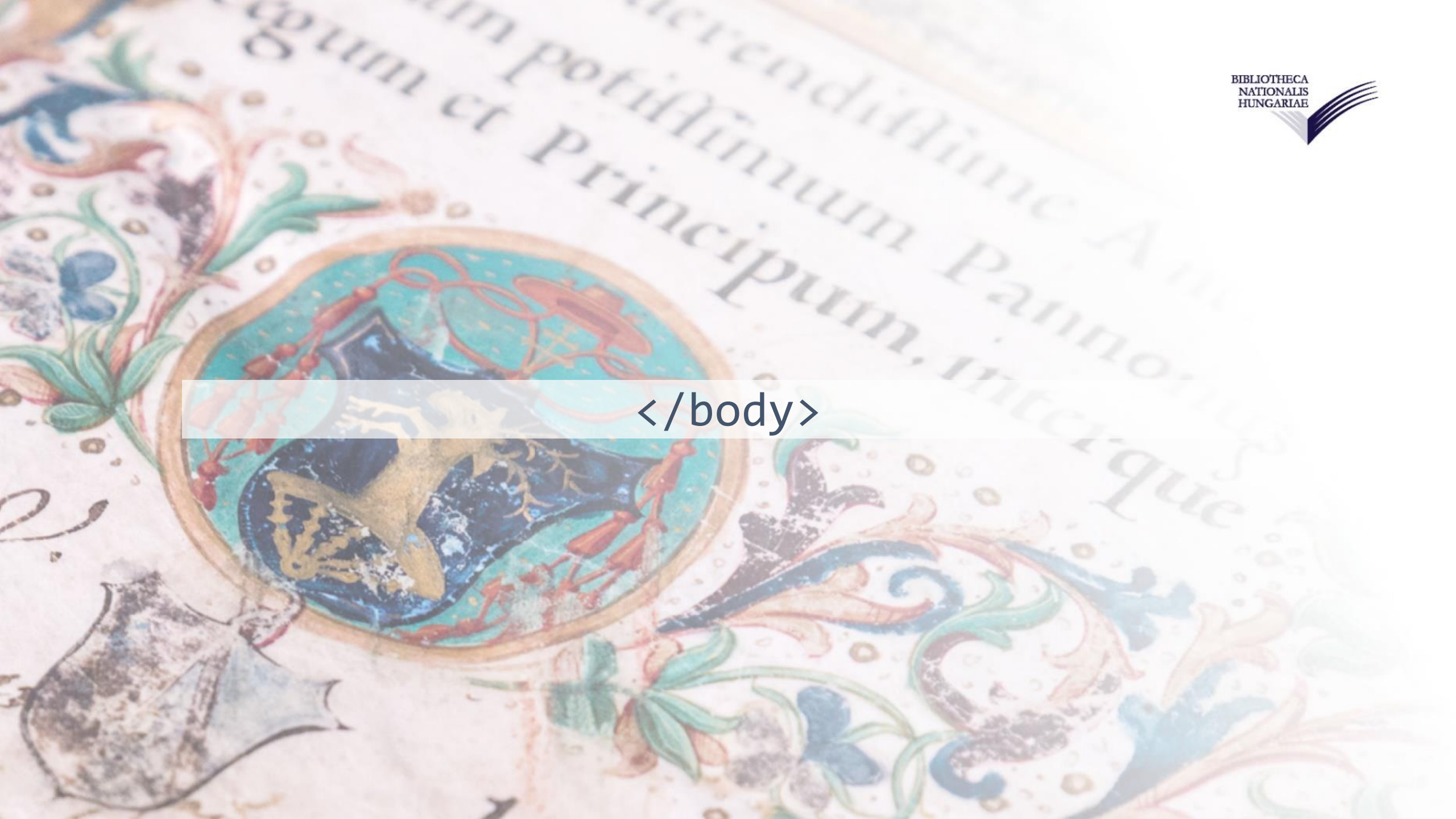


# RFC5322 szerinti header fiek

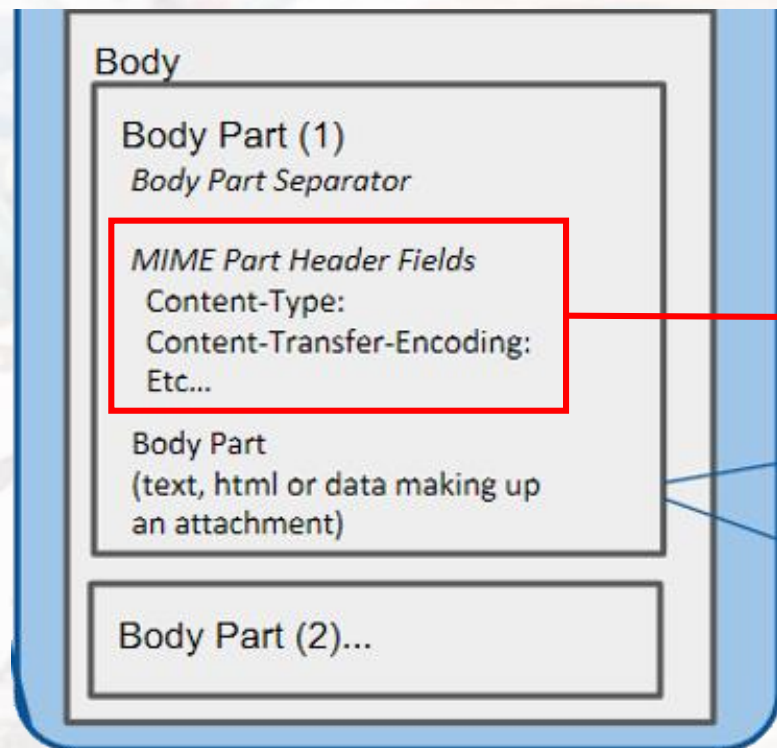
Dublin Core Metadata Element Set/Simple Dublin Core	
<b>Contributor</b>	An entity responsible for making contributions to the resource.
<b>Coverage</b>	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
<b>Creator</b>	An entity primarily responsible for making the resource.
<b>Date</b>	A point or period of time associated with an event in the lifecycle of the resource.
<b>Description</b>	An account of the resource.
<b>Format</b>	The file format, physical medium, or dimensions of the resource.
<b>Identifier</b>	An unambiguous reference to the resource within a given context.
<b>Language</b>	A language of the resource.
<b>Publisher</b>	An entity responsible for making the resource available.
<b>Relation</b>	A related resource.
<b>Rights</b>	Information about rights held in and over the resource.
<b>Source</b>	A related resource from which the described resource is derived.
<b>Subject</b>	The topic of the resource.
<b>Title</b>	A name given to the resource.
<b>Type</b>	The nature or genre of the resource.

- ← dátumra vonatkozó
- ← feladóra vonatkozó
- ← címzetre vonatkozó
- ← identifikációs
- ← információ
- ← továbbított üzenet header mezői

</body>







Típus/Altípus	Leírás
text/plain	ASCII szöveg (ez az alapértelmezett tartalomtípus)
text/html	HTML fájl
text/javascript	Javascript kód
text/CSS	CSS fájl
image/gif	GIF kép
image/jpeg	JPEG kép
audio/mp3	MP3 zene
audio/x-wav	WAV hangfájl
video/mpeg	MPEG film
application/octet-stream	közelebbről meg nem határozott jelentésű bájt sorozat
application/javascript	JavaScript fájl
application/json	JSON fájl
application/vnd.ms-excel	Microsoft Excel dokumentum
application/x-shockwave-flash	Adobe Flash fájl
multipart/mixed	többrészes üzenet
multipart/alternative	egy üzenet több különböző formátumban



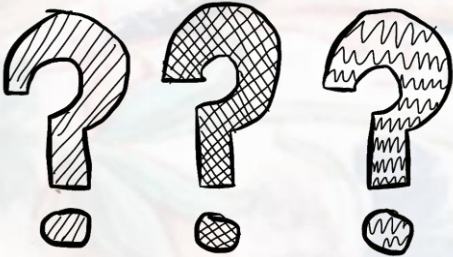


# Csomagolás – MailBag



```
<base directory>/
|
+-- bagit.txt
|
+-- bag-info.txt
|
+-- mailbag.csv
|
+-- manifest-<algorithm>.txt
|
+-- tagmanifest-<algorithm>.txt
|
+-- data/
|
+-- mbox/
|   +-- [payload files]
+-- pdf/
|   +-- [payload files]
+-- warc/
|   +-- [payload files]
+-- attachments/
|   +-- [Mailbag-Message-ID]/
|   |   +-- [payload files]
|   +-- [Mailbag-Message-ID]/
|   |   +-- [payload files]
|   ...
|
...
```

```
<base directory>/
|
+-- bagit.txt
|
+-- bag-info.txt
|
+-- manifest-<algorithm>.txt
|
+-- tagmanifest-<algorithm>.txt
|
+-- <Message-ID>_metadata.txt
|
+--... [mindegyik e-mailhez külön a metadata.txt]
|
+--... MBOX_Bitcurator_report_1.txt
|
+--... [az összes report]
|
+-- data/
|
+-- mbox/
|   +-- [payload files]
+-- eml/
|   +-- [payload files]
+-- txt/
|   +-- [payload files]
+-- html/
|   +-- [payload files]
+--?/
|   +-- [payload files]
+-- attachments/
|   +-- [Mailbag-Message-ID]/
|   |   +-- [payload files]
|   +-- [Mailbag-Message-ID]/
|   |   +-- [payload files]
|   ...
|
...
```





**Thank you for your attention!**

**Time for Whiskas!**

Köszönöm a figyelmet!