

Academia Română – Biblioteca  
București, Sector 1, Calea Victoriei nr. 125  
Telefon : (021)212.82.84  
Fax : (021)212.58.56  
Adresă de web: www.biblacad.ro

## **RAPORT DE DEPLASARE** **HAGA 29 MAI – 1 IUNIE 2011**

### **A. CADRU :**

În perioada 30 și 31 mai anul curent a avut loc la Biblioteca Regală a Olandei din Haga (KB), în cadrul Proiectului Bibliotecile Europeana, reuniunea comună a următoarelor grupuri:

1. TEL Library Coordination Group (Grupul TEL pentru Coordonarea Bibliotecilor)
2. TEL Marketing and Editorial Working Group (Grupul TEL de Lucru pentru Marketing și Editare)
3. Europeana Libraries Metadata Working Group (Grupul Bibliotecile Europeana de Lucru cu Metadatele)
4. Europeana Libraries Full Text Content Provider (Grupul Furnizorilor de Conținut în baze Full Text din Bibliotecile Europeana)



### **B. Reprezentarea Bibliotecii Academiei Române:**

Biblioteca Academiei Române a fost prezentată de o delegație formată din două persoane, ambele membre ale echipei care lucrează în cadrul acestui proiect european:

1. **Flori Dumitrescu**, informaticianul echipei de proiect a Bibliotecii Academiei Române
2. **Ligia Caranfil**, bibliograf și secretarul echipei de proiect a Bibliotecii Academiei Române

### **C. Durata șederii:**

Delegația Bibliotecii Academiei Române a sosit la Haga pe data de 29 mai anul curent și a plecat spre București în data de 1 iunie anul curent.

### **D. Desfășurare:**

Întrunirea s-a desfășurat conform programului care ne-a fost comunicat de către organizatori.

### **D<sub>1</sub>. Prima zi, luni 30 mai 2011:**

10:00-2:15

ședința plenară, în aula Bibliotecii Regale a Olandei.

În cadrul acestei ședințe au luat cuvântul:



1. **Arianna Ciula**, Science Officer Humanities, de la European Science Foundation (Fundatia Europeana a Științelor), – *Research Infrastructures In the Digital Humanities* (Infrastructuri de Cercetare în Științele Umaniste digitale). În cadrul acestui discurs a făcut o prezentare Fundației Europene a Științelor atât din punct de vedere instituțional, cât și al activității sale, dar și a infrastructurilor de cercetare. Discursul doamnei Ciula a inclus și o informație foarte importantă, în legătură cu conservarea și sustenabilitatea patrimoniului și anume că doar 22 de instituții de patrimoniu cultural din Europa și-au făcut la ora actuală un plan de conservare pe termen lung.
2. **Louise Edwards**, de la European Library (Biblioteca Europeana), care a prezentat foarte pe scurt proiectul European Libraries și primul pachet de lucru din cadrul proiectului, al cărui lider este.
3. Fiecare dintre liderii celorlalte cinci pachete de lucru au trecut la microfon și au prezentat foarte pe scurt pachetele de lucru.

Doamna Susan Reilly, liderul pachetelor de lucru 2 si 4 a făcut precizarea că va urma să fie distribuit spre completare și semnare un chestionar similar ca cel pentru pachetul de lucru 3, care a fost discutat a doua zi, în cadrul workshop-ului *Data Aggregation Clinic*.

12:15-13:15

pauză de prânz în foaierul aulei.

13:15-13:45

cele 2 ședințe pe grupe. Delegația BAR a fost repartizată la Europeana Libraries & Technical and Interoperability Working Group (Bibliotecile Europeana & Grupul de Lucru Tehnic și pentru Interoperabilitate). În cadrul acestei sesiuni, care s-a ținut tot în aulă, au luat pe rând cuvântul:

- **Jose Borbinha** a propus scindarea grupului in trei grupulețe mai mici: furnizorii de conținut actuali, noii furnizori de conținut si viitorii furnizori de conținut. Noi am intrat în a doua subgrupă, cea a noilor furnizori de conținut. Domnul Borbinha a specificat si că nu există în prezent un model definit pentru livrarea de full-text către Europeana si tocmai acesta este subiectul workshop-ului dedicat bazelor full-text. Domnul Borbinha a împărțit tuturor textul în ciornă al primului livrabil din pachetul de lucru 4, intitulat: *Requirements Infrastructure and Harvester* (Cerințe pentru Infrastructura de Recoltare a Datelor și Serviciul de Recoltare a Datelor), cu rugămintea să-l citim și să-i trimitem un *feedback* în decurs de 2 săptămâni.
- **Sally Chambers** - *From full-text discovery to textual analysis: an exploration of user needs for full-text (De la descoperirea full-text-ului la analiza textuală: o explorare a necesităților utilizatorilor pentru full-text)*. În prezentarea sa, Sally Chambers arată că unul dintre obiectivele centrale ale proiectului Bibliotecile Europeana este agregarea bazelor full-text ale bibliotecilor de cercetare din Europa. Pornind de la problema reinventării cercetării științifice, doamna Chambers propune un portal de agregare. Atenția doamnei se centrează pe științele umaniste, la fel ca si in prezentarea doamnei Ciula. Un obiectiv central stabilit de domnia sa este ușurința și viteza accesului la resursele digitale. Doamna Chambers a făcut portretul-robot al unei biblioteci digitale ideale cu un conținut alcătuit din: toate textele existente într-o bibliotecă obișnuită, fie ele manuscrise sau tipărite, toate periodicele de orice tip, tot ceea ce înseamnă arte vizuale etc. Toate acestea trebuie prezentate atât ca full-text cât și ca full-image. Toate acestea să fie catalogate în mod adecvat. Accesul la colecțiile unei asemenea biblioteci digitale trebuie să dea posibilitatea cercetării bazelor



full-text, accesul să fie nemijlocit, nu prin portaluri locale la colecții specifice, să nu se facă doar prin interfață web, obiectele digitale să poată fi accesate în diverse formate: text, web etc., biblioteca să poată oferi diverse servicii de facilitare a căutării, descoperii de conținut. Printre instrumentele de regăsire pe care o bibliotecă digitală ideală ar trebui să le pună la îndemâna cercetătorilor săi ar fi: indicii de recunoaștere a entităților digitale, traduceri, dicționare, adnotări etc. Sally Chambers a dat și un exemplu de full-text, este vorba de *Old Baily Online*, care se găsește la adresa: <http://www.oldbaileyonline.org/>

The image displays two screenshots of the Old Bailey Online website, showing the homepage and search interface. The browser used is Mozilla Firefox.

**Top Screenshot:**

- Page Title: Old Bailey Online - The Proceedings of the Old Bailey, 1674-1913 - Central Criminal Court - Mozilla Firefox
- Address Bar: <http://www.oldbaileyonline.org/>
- Navigation Menu: Home | Search | About The Proceedings | Historical Background | The Project | Contact
- Section: **The Proceedings of the Old Bailey, 1674-1913**
- Description: A fully searchable edition of the largest body of texts detailing the lives of non-elite people ever
- Search Box: SEARCH the Proceedings, Keyword(s), Reference No.
- Left Sidebar: Home Page, Search, About the Proceedings, Historical Background, The Project

**Bottom Screenshot:**

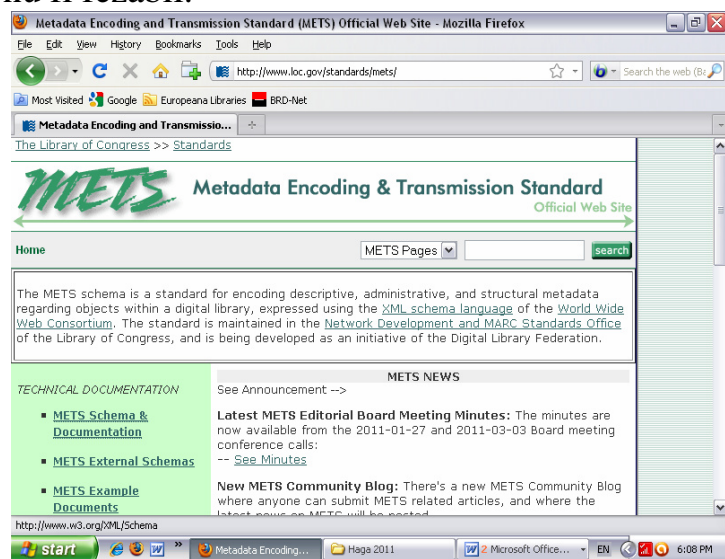
- Page Title: Old Bailey Online - The Proceedings of the Old Bailey, 1674-1913 - Central Criminal Court - Mozilla Firefox
- Address Bar: <http://www.oldbaileyonline.org/>
- Navigation Menu: Home Page, Search, About the Proceedings, Historical Background, The Project, Copyright & Citation Guide, Contact, Research and Study Guides, Login / Register
- Section: **The Proceedings of the Old Bailey, 1674-1913**
- Description: A fully searchable edition of the largest body of texts detailing the lives of non-elite people ever published, containing 197,745 criminal trials held at London's central criminal court. If you are new to this site, you may find the [Getting Started](#) and [Guide to Searching](#) videos and tutorials helpful.
- Search Box: SEARCH the Proceedings, Keyword(s), Reference No., Search In: <All Text>, SEARCH, More Search Options
- Left Sidebar: Home Page, Search, About the Proceedings, Historical Background, The Project, Copyright & Citation Guide, Contact, Research and Study Guides, Login / Register
- Special Section: **ON THIS DAY IN... 1854**  
Charles Ward killed a sixteen month old baby while driving under the influence of alcohol. [read more](#)
- Footer: To search the Proceedings use the boxes on the right or go to **CENTRAL CRIMINAL COURT. NEWTON'S PAPER.**

- **Gilbert Pedrosa**, de la Instituto Superior Tecnico din Portugalia, a prezentat proiectul TELPlus, legat de portalul TEL. Obiectivul celor de la TEL este să găsească o soluție pentru organizarea unui serviciu central de indexare a bazelor full-text în TEL. Au folosit sistemul SOLr pentru indexarea tuturor datelor și informațiilor. Vorbitorul a insistat pe nevoia unui Manager Unificat de Ingerare a Datelor. Rezultatele activității lor din cadrul proiectului TELPlus sunt următoarele:

- Noi elemente de metadate pentru TEL-AP
- Trei mecanisme de recoltare a full-text-ului.
- Dezvoltarea unor programe de calculator speciale (software)
- Motor de indexare bazat pe SOLr
- Integrarea cu portalul TEL folosind interfața USR
- Analiza inițială a primelor seturi de date este adusă la cunoștința publicului
- Metode de acces
- Formate de fișiere
- Utilizarea metadatelor
- Link-area full-text-ului
- Link-area metadatelor structurale

Totodată, domnul Pedrosa a criticat formatul METS (Metadata Encoding & Transmission Standard):

- Doar o singură bibliotecă și-a livrat metadatele în acest format
- A prezentat mai multe dificultăți și s-a dovedit a nu fi fezabil.



În continuare, acest vorbitor a subliniat importanța integrării datelor în portalul TEL și pe integrarea ingerării datelor. Bazele full-text au fost integrate ca o temă separată.

- **Clemens Neudecker**, de la Biblioteca Regală a Olandei, care lucrează la proiectul IMPACT - *Technical solutions and future challenges for full-text content, experiences from IMPACT (Soluții tehnice și provocări viitoare pentru conținutul de full-text, experiențe ale proiectului IMPACT)*. A prezentat foarte pe scurt proiectul IMPACT:

- Proiect FP7 finanțat de Comisia Europeană
- Coordonat de KB
- 26 de parteneri: biblioteci, institute de cercetare, parteneri din industrie
- Obiectivul central este îmbunătățirea accesului la text.
- Obiectivele acestui proiect sunt: îmbunătățirea ocr-ului, dezvoltarea unor instrumente diferite, crearea unor dicționare computerizate pentru 9 limbi vechi diferite, digitizarea să fie mai rapidă, mai bună, mai ieftină.
- Principala țintă a acestui proiect sunt limbile în dezvoltarea lor diacronică, istorică.
- Toate soluțiile IMPACT sunt programe de calculator (*software*).
- IMPACT s-a concentrat pe corecția și armonizarea lingvistică și pe textul tipărit.
- Oferă recunoașterea entităților după:
  - Nume persoană,
  - Nume geografic
  - Nume organizație
- Oferă un cadru de interoperabilitate
- Oferă o serie de instrumente strategice:
  - Recomandări pentru cele mai bune practici
  - Helpdesk
  - Studii de caz
  - Instrumente pentru planificarea digitizării.

Adresa din cadrul site-ului IMPACT unde se pot găsi astfel de instrumente este: <http://www.impact-project.eu/taa/strat/pilot-tools/>

Impact | Improving access to text: Pilot Tools - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.impact-project.eu/taa/strat/pilot-tools/

Most Visited Google Europeana Libraries BRD-Net

Old Bailey Online - The Pr... The European Library Google Traducere Impact | Improving a...

Improving Access to Text  
**IMPACT**

printable view "Search the IMPACT website"

Home  
News  
Helpdesk

Tools and applications  
Calendar of events  
About the project  
FAQs  
Documents  
Sitemap  
Disclaimer  
Contact  
For partners

Demonstrator Platform  
Workflow Client  
Web Service Client  
CONCERT  
Metrics Toolkit  
Technical  
Tools  
Strategic  
Tools  
Pilot Tools

**IMPACT PILOT-RELEASE MATERIALS**

**Metadata for Text Digitisation & OCR**  
[IMPACT-metadata-bpg-pilot-1.pdf](#)  
 IMPACT Best Practice Guide: Metadata for Text Digitisation & OCR 671 K

**OCR for Mass Digitisation**  
[IMPACT-ocr-bp-pilot-1b.pdf](#) 148 K  
[IMPACT Briefing Paper: OCR](#)  
[IMPACT-ocr-bpg-pilot-s1.pdf](#) 253 K  
[IMPACT Best Practice Guide: OCR Section 1](#)  
[IMPACT-ocr-bpg-pilot-s2.pdf](#) 820 K  
[IMPACT Best Practice Guide: OCR Section 2](#)  
[IMPACT-ocr-bpg-pilot-s3.pdf](#) 157 K  
[IMPACT Best Practice Guide: OCR Section 3](#)

Done

start Impact | Improving a... Microsoft Office ... EN 5:17 PM

Impact | Improving access to text: Pilot Tools - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.impact-project.eu/taa/strat/pilot-tools/

Most Visited Google Europeana Libraries BRD-Net

Old Bailey Online - The Pr... The European Library Google Traducere Impact | Improving a...

WordPress  
slideshare  
YouTube vimeo

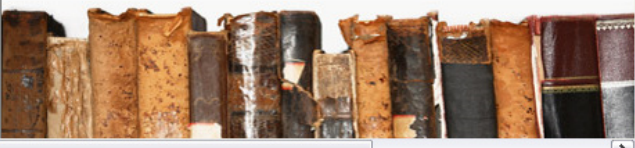
Login:  
username  
password  
enter

[impact\\_case-study\\_digitising-from-microfilm.pdf](#)  
 IMPACT Case Study - Digitising from Microfilm 1.7 M

**IMPACT PILOT-RELEASE SUPPORT TOOLS**

**IMPACT Storage Estimator (ISE)**  
[IMPACT-ise-4.xls](#)  
 IMPACT Storage Estimator V4 451 K  
[IMPACT-ise-tutorial-4.pdf](#)  
 IMPACT Storage Estimator V4 - Tutorial 1.2 M

**IMPACT Digitisation Cost Estimator (IDCE)**  
[IMPACT-idce-3a.xls](#)  
 IMPACT Digitisation Cost Estimator V3 134 K



Done

start Impact | Improving a... Microsoft Office ... EN 5:18 PM



Microsoft Excel - IMPACT-idce-3a [Read-Only]

File Edit View Insert Format Tools Data Window Help

Arial 10


K2

### 1. DEFINE BASIC PARAMETERS

# of volumes	3,000
# of total pages to be imaged	4,500,000
storage space per page incl. surrogates (Mb.)	40
cost of storage per terabyte	500 €
working days in a year for FTE	220
working hours in a day for FTE	8
QA overhead	15%

This tool will estimate the overall cost of undertaking a digitisation project. To use it, please fill out the light green cells in Section 1 and then Section 2.

This tool is released under a Creative Commons License: Attribution-NonCommercial-ShareAlike 3.0 Unported



### 2. SET YOUR WORKFLOW STEPS

	Step 1	Step 2	Step 3	Step 4	Step 5	Workflow totals
# of minutes per volume	3	1			1	
# of minutes per 10 pages		1.5	1			3
# of personnel (FTE)	4	4	5			13
annual gross salary per FTE	42,000 €	18,000 €	36,000 €		24,000 €	120,000 €
time cost in days (incl. QA)	5	406	216	0	543	1,170
overall personnel costs	4,116 €	132,903 €	176,420 €	0 €	118,398 €	431,838 €
outsourcing / 3rd party costs				150,000 €		150,000 €

### 3. CALCULATION REPORT

total man-days in project	1,170
total project duration in days	548
total project duration in years	2.49
total storage req. in Tb.	180

We hope that the use of this tool will be self-explanatory, however, if you would like any further support in its use, please contact the IMPACT HelpDesk at [www.impact-project.eu/helpdesk](http://www.impact-project.eu/helpdesk).

Although great care was taken in producing this cost estimator, you use it at your own risk. IMPACT can not guarantee the absence of errors or omissions in this tool. If you find any mistakes, please help us improve the tool by either leaving feedback with the HelpDesk or by contacting the author at [impact@sub.uni-goettingen.de](mailto:impact@sub.uni-goettingen.de)

Digitisation Cost Estimator /

Ready

start Impact | Improvi... Microsoft Off... Microsoft Excel -... EN 5:26 PM


## IMPACT: costuri estimative

Microsoft Excel - IMPACT-idce-3a [Read-Only]

File Edit View Insert Format Tools Data Window Help

Arial 10

E25



### 2. SET YOUR WORKFLOW STEPS

	Step 1	Step 2	Step 3	Step 4	Step 5	Workflow totals
# of minutes per volume	3	1			1	
# of minutes per 10 pages		1.5	1			3
# of personnel (FTE)	4	4	5			13
annual gross salary per FTE	42,000 €	18,000 €	36,000 €		24,000 €	120,000 €
time cost in days (incl. QA)	5	406	216	0	543	1,170
overall personnel costs	4,116 €	132,903 €	176,420 €	0 €	118,398 €	431,838 €
outsourcing / 3rd party costs				150,000 €		150,000 €

We hope that the use of this tool will be self-explanatory, however, if you would like any further support in its use, please contact the IMPACT HelpDesk at [www.impact-project.eu/helpdesk](http://www.impact-project.eu/helpdesk).

Although great care was taken in producing this cost estimator, you use it at your own risk. IMPACT can not guarantee the absence of errors or omissions in this tool. If you find any mistakes, please help us improve the tool by either leaving feedback with the HelpDesk or by contacting the author at [impact@sub.uni-goettingen.de](mailto:impact@sub.uni-goettingen.de)

Digitisation Cost Estimator /

Ready

start Impact | Improvi... Microsoft Off... Microsoft Excel -... EN 5:28 PM



Domnul Neudecker a discutat provocările procesului de ocr-izare. IMPACT propune binarizarea în cazul ocri-izării textelor vechi cu caractere neclare, iar pentru adnotările pe texte, binarizarea adaptativă. Mai propun eliminarea marginilor paginilor, corectarea geometrică, pentru paginile întoarse sau mototolite, sau Dewarping, care mai este potrivită și pentru materialele multicolore. Unul dintre rezultatele acestui proiect a fost îmbunătățirea recunoașterii unor caractere mai speciale cum ar fi s lung (scharfes S), caracterele gotice etc. Pentru textele tipărite prost, cu litere șterse, neclare etc., IMPACT propune o corectură în cooperare. Au reușit îmbogățirea rezultatelor ocr-izării cu informații structurale. În ceea ce privește limba din perspectivă diacronică și diversele ortografii ale limbilor în istoria lor, IMPACT a elaborat dicționare speciale pentru 9 limbi: olandeză, germană, engleză, franceză, spaniolă, poloneză, slovenă, bulgară și cehă.

La discuția care a urmat după prezentarea domnului Neudecker, am aflat că Rusia și Grecia dezvoltă fiecare în parte programe de ocr-izare a manuscriselor.

16:00-17:30

workshop-ul care s-a desfășurat în sala B și a fost prezidat de:

1. **Jose Borbinha**, de la *Instituto Superior Técnico*,
2. **Gilberto Pedrosa**, de la *Instituto Superior Técnico*,
3. **Ina Heinz** de la TEL.

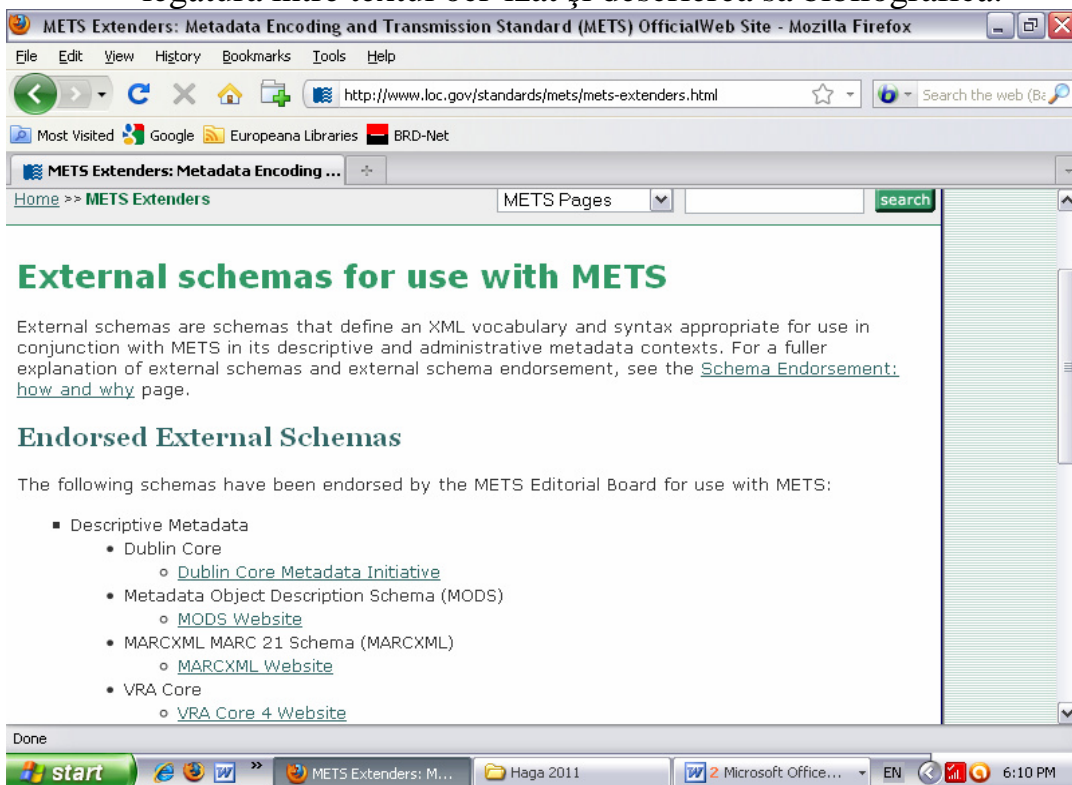
A fost o discuție liberă, la care participanții, toți furnizorii de conținut, și-au prezentat foarte pe scurt conținutul pe care doresc să-l livreze către Europeana și au putut pune întrebări celor de la TEL. Ni s-au pus la toți următoarele întrebări:

1. Ce tehnologii folosesc furnizorii de conținut pe plan local pentru crearea de full-text?
2. Ce planuri referitoare la full-text au furnizorii de conținut, dacă nu am început deja să lucrăm la full-text?
3. Care sunt condițiile elementare pe care trebuie să le îndeplinească furnizorii de conținut?
4. Domnul Borbinha a rugat-o pe doamna Flori Dumitrescu să-i arate câteva entități digitizate la Biblioteca Academiei Române.

Din aceste dezbateri interactive s-au desprins următoarele idei:

- Definiția full-text-ului în contextul proiectului Bibliotecile Europeana: metadată structurală, obiect digital ocr-izat. **Să se observe că procesul de ocr-izare intră în însăși definiția conceptului de full-text. Full-textul, atunci nu poate fi conceput în afara ocr-izării.**
- Problema format-ului este irelevantă.

- Programul METS poate fi distribuit gratuit partenerilor din proiect care au nevoie de un software care să construiască legătura între textul ocr-izat și descrierea sa bibliografică.



- Conceptul de imagine transparentă este necesar pentru formatul pdf.
- Conceptul de URL persistent, o soluție la eventualele dificultăți apărute în urma schimbărilor survenite în structura metadatelor (schimbări legate de format sau de sistemul integrat de bibliotecă etc.).
- Conceptul de *tiny* URL (URL minuscul) care este același lucru cu cel de URL persistent. O bibliotecă își poate crea propriul serviciu de URL-uri minuscule.
- Cei de la TEL preferă de departe formatele MARC, fie ele UNIMARC, MARC 21 etc., altor tipuri de formate, precum DublinCore, pentru că sunt cele mai bogate tipuri de formate.
- Cei de la TEL păstrează formatul original al metadatelor importate.
- Furnizorii de conținut își exportă datele așa cum le au, în formatele în care le au, iar agregatorul va face tot ceea ce ține de prelucrarea datelor, de maparea lor, de indexarea lor, de ingestia și de agregarea lor.
- Nu este în intenția echipei TEL să dezvolte un catalog unificat, ci un catalog pentru descoperiri, o schemă foarte simplă, bazată pe mapări.
- Gilberto Pedrosa a propus un scenariu de viitor, în doi pași:

1. Recoltarea metadatelor
  2. Recoltarea ocr-urilor
- Plan de ingestie a colecțiilor în primul trimestru al acestui an, doar pentru metadata.
  - Problema vocabularului controlat în contextul proiectului Bibliotecile Europeana. Nu există nici o preocupare pentru vocabular controlat în cadrul acestui proiect.
  - Principala problemă legată de full-text este structurarea conținutului cu full-text-ul. Un posibil răspuns la această problemă sunt identificatorii persistenți.
  - Problema unei limite temporare pentru ocr-izarea unui text. Cei de la TEL ne-au asigurat că nu există o asemenea limită, dar ne-au pus în vedere următorul fapt: cu cât este textul mai vechi, cu atât problemele și, respectiv, cheltuielile de ocr-izare sunt mai mari
  - Problema duplicatelor. Se dorește versiuni diferite ale aceluiași obiect digital provenind de la o imagine originală cu html.

19:30

Cina festivă la Rootz.

## **D<sub>2</sub>. A doua zi, marti, 31 mai 2011:**

10:00-10:15

ședință plenară în aula Bibliotecii Regale a Olandei, în care s-a făcut o foarte scurtă prezentare a programului acelei zile.

10:15-11:00

Introducere în Modelul de Date Europeana (EDM).

Au luat cuvântul:

1. **Sally Chambers**, de la TEL: *Introduction to Europeana Data Model (EDM) (Introducere la Modelul de Date Europeana (EDM))*, care a început prin a preciza obiectivele acestei prezentări:
  - a. Facilitarea căutării conținutului existent din domeniul bibliotecilor prin alinierea metadatelor din domeniul bibliotecilor la EDM.
  - b. Înțelegerea EDM.
  - c. Explorarea posibilităților de aplicare ale EDM la metadatale de bibliotecă.
  - d. Maparea metadatelor din domeniul bibliotecilor la EDM.
2. **Robina Clayphan**, de la TEL: *EDM Rationale and Overview of Classes and Properties (Principiile EDM și o trecere în revistă a claselor și proprietăților)*. Doamna Clayphan a pornit de la ESE,

Elementele Semantice Europeana, numitorul comun al metadatelor obiectelor digitale în Europeana și le-a caracterizat arătându-le mai ales dezavantajele:

- a. Forțează interoperabilitatea,
- b. Model plat, alcătuit mai ales din valori ale șirurilor textuale,
- c. Violarea principiului unu-la-unu,
- d. Pierderea bogăției metadatelor originale, fiind principalul neajuns al ESE.

La ora actuală, ESE constituie un subset al EDM, iar furnizorii de conținut pot continua să suplimenteze ESE. Apoi a enumerat cerințele EDM:

- a) Să distingă între obiectul din realitate și reprezentarea lui digitală,
- b) Să distingă între obiect și metadatele care îl înregistrează și îl descriu,
- c) Să permită mai multe înregistrări diferite pentru același obiect,
- d) Să includă obiecte compuse din alte obiecte,
- e) Să ofere un format standard pentru metadate care să poată fi specializat,
- f) Să ofere un vocabular standard pentru metadate care să poată fi specializat,
- g) Să se bazeze pe standardele existente.

După aceea a vorbit despre principiile fundamentale ale EDM. Imaginea de bază a acestui model este:

### **rețea semantică + obiecte în rețea.**

Prin definirea obiectelor se ajunge la clasele EDM. Un exemplu de clasă EDM care ar interesa bibliotecarii este clasa resurselor noninformaționale:

- Eveniment
- Agent
- Loc
- Entitate fizică
- Concept
- Perioadă de timp.

Proprietățile EDM descriu clasele EDM. Trebuie să existe *link*-uri între proprietăți și clase. Agregarea organizează datele unui singur furnizor. Pentru doi furnizori ai aceluiași obiect se vor efectua două agregări separate. Crearea datelor EDM se face aplicând diferitelor obiecte diverse proprietăți din sfera modelului. Vor exista linii directoare care să ajute la crearea datelor EDM. Din punctul de vedere al echipei de la Biblioteca Europeană, care se ocupă de dezvoltarea acestui model, furnizorii de conținut nu au nevoie să aplice direct modelul la obiectele



lor, ci doar să furnizeze maparea pentru aceste obiecte, care vor migra către Modelul de Date Europeană. [Ceii de la TEL și-au luat asupra lor sarcina mapării datelor recoltate de la furnizorii de conținut.](#) Există mapări inițiale de la datele existente în ESE către EDM. Ce le mai rămâne de făcut celor de la Biblioteca Europeană în cazul EDM:

- Finalizarea definițiilor conform cărora fiecare proprietate va fi atribuită câte unei clase
- Finalizarea schemei XML
- Finalizarea acelor părți din EDM care urmează să fie primele implementate.

3. **Valentine Charles**, de la TEL/Europeană: *Object Templates, Applying Properties to Classes (Machetele obiectelor digitale, aplicarea proprietăților la clase)*. A arătat că acest model este structurat pe clase. A definit conceptul de clasă: un grup de entități care au în comun aceleași proprietăți. Proprietățile sunt atașate claselor. A definit conceptul de proprietate: un atribut sau o caracteristică a unei resurse. Apoi a vorbit despre principalele noțiuni ale EDM:

- a. Obiect
- b. Reprezentare digitală, definită ca forma digitală a unui obiect real însoțit de o adresă de web.

Cele două reunite formează o agregare. A definit conceptul de agregare: organizează informațiile legate de un obiect provenite de la un furnizor de conținut particular. Proprietățile atribuite unui obiect sunt atribute ale obiectului din realitate, nu ale reprezentării lui digitale. După Valentine Charles, EDM mai oferă următoarele avantaje:

- Relații între obiectele furnizate
- Link-uri tip parte-întreg pentru obiecte compuse (ierarhice)
- Relații derivative și între diverse versiuni digitale
- Resurse informaționale care dispun de cel puțin o reprezentare web și cel puțin un URI
- Proprietăți pentru agregare
- Clasele sunt astfel structurate încât să ofere informații contextuale cu privire la obiecte
- Aceste clase sunt folosite la producerea unor descrieri îmbogățite, mai ales atunci când există vocabulare controlate, folosite de instituțiile furnizoare de conținut.
- Resurse contextuale

- Informații contextuale

Valentine Charles a mai arătat și că setul de resurse referitoare la un singur obiect de patrimoniu cultural, ce reprezintă în mod colectiv acel obiect în Europeana, constă din totalitatea descrierilor privitoare la obiectul pe care Europeana îl colectează de la furnizorii de conținut (posibil diferiți), inclusiv thumbnails și alte forme de abstracție.

11:45-13:00

Content Aggregation Clinic (Clinica de Agregare a Datelor), în sala B. În cadrul acestui workshop s-a discutat Planul Bibliotecii Europene de Ingestie a Conținutului. Discuțiile au fost prezidate de echipa de agregare de la TEL:

1. **Magali Mermet** de la TEL,
2. **Chiara Latronico** de la TEL
3. **Susanna Summa** de la Europeana/TEL

În acest context a fost discutat chestionarul pentru pachetul de lucru 3, completat și semnat de Biblioteca Academiei Române. La acest workshop am aflat următoarele lucruri:

- Conform acestui chestionar, numărul total de itemi se ridică la 2.138.971, din care Bibliotecii Academiei Române îi revine un număr de 3.440.
- S-a făcut un sumar al rezultatelor preconizate.
- S-a trecut la o demonstrație de ingestie.
- Recoltarea datelor se face via FTP.
- Vor începe recoltarea din luna iulie a acestui an.
- Planul de Ingestie a Conținutului ne-a fost prezentat defalcat pe patru trimestre din an, în care erau programate recoltările de la bibliotecile membre ale proiectului Bibliotecile Europeana. Noi intrăm abia în al treilea trimestru de recoltare, cu alte cuvinte, abia în luna ianuarie a anului viitor. Principalele provocări ale acestui plan pentru furnizorii de conținut sunt:
  - a) respectarea termenului limită,
  - b) exportarea la timp a conținutului,
  - c) dacă noi suntem gata și ei vor fi gata.
- Când ne va veni rândul la recoltarea datelor, echipa TEL ne va contacta printr-un reminder email cu o lună înainte de scadență, vom face împreună o probă de recoltare, apoi ni se va cere un *feedback*.
- Echipa TEL este interesată de date bibliografice, nu și agregatorul Europeana.
- Este mai bine să se separe în două seturi obiectul digital de descrierea lui bibliografică. Europeana își va crea propriile *thumbnails*.

- S-a pus problema limbii în care vor fi recoltate metadatele. Dacă metadatele vor fi prezentate în două sau mai multe limbi, vor fi preluate ca atare.

13:00

Pauză de prânz, în foaietul aulei.

14:00-15:30

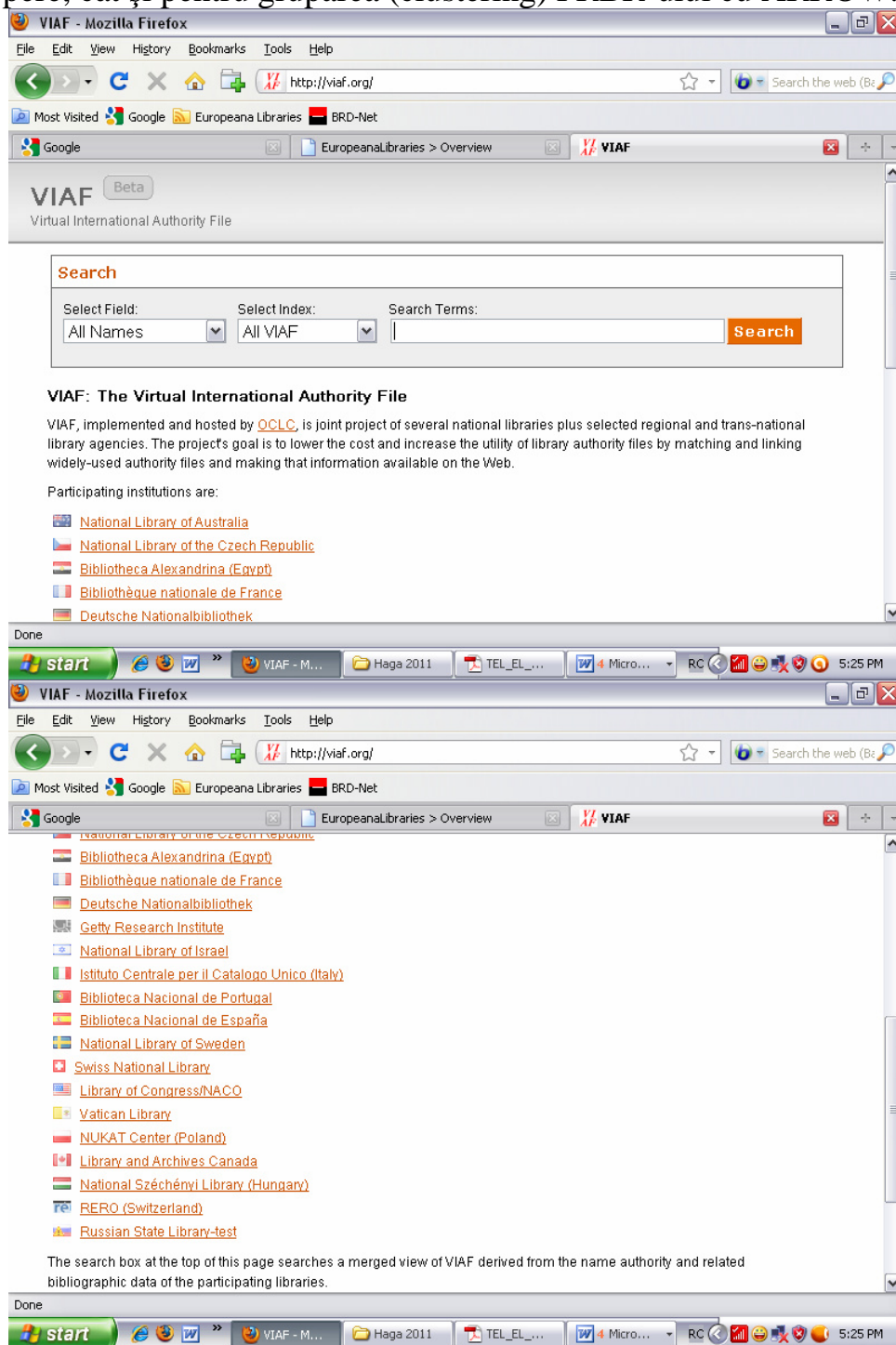
Data Mining (Prelucrarea datelor). Acest workshop, care s-a desfășurat tot în sala B, a fost prezidat de **Andreas Juffinger** de la TEL.



Din acest workshop am cules următoarele informații:

- Conceptul de *data mining*, prelucrarea propriu-zisă a datelor, presupune următoarele elemente centrale:
  - Datul
  - Fațeta. Se pune întrebarea: care fațetă este într-adevăr relevantă pentru *end users*?
  - *Cluster*-ul, activitatea de *clustering*, de grupare a datelor și de formare a grupurilor sau configurațiilor de date.
- Scurtă prezentare introductivă a proiectului ARROW (Accessible Registries of Rights Information and Orphan Works – Registre Accesibile de Informații privitoare la Drepturile de Autor și la Operele Orfane). Prezentarea fluxului de lucru în cadrul acestui proiect, care permite utilizatorilor să verifice, prin intermediul unei interfețe dezvoltate la nivel european, dacă o lucrare este disponibilă, expirată sau orfană și să obțină informații privind deținătorii drepturilor. Acest proiect este legat, mai ales, de digitizarea cărților în biblioteci.
- VIAF (the Virtual International Authority File – Fișierul Virtual Internațional de Autoritate) este relevant pentru FRBR(Functional

Requirements for Bibliographic Records – Cerințe Funcționale pentru Înregistrări Bibliografice), atât fiindcă deține date despre autorii de opere, cât și pentru gruparea (clustering) FRBR-ului cu ARROW.



- Statutul curent al susținerii TEL pentru VIAF: s-a implementat deja interfața cu serviciul web al VIAF și este implicată în activitățile productive ale sistemului TEL/ARROW.
- Ingestia cataloagelor de bibliotecă către TEL.
- Mecanisme de transfer
- Problema reduplicării conținutului



- Problema fațetării datelor. Fațetarea este definită în acest context ca o modalitate suplimentară de grupare/configurare (*clustering*) a datelor. Definiția fațetei.
- Problema limbii: limba metadatelor sau limba obiectului?
- Problema stabilirii vedetei de subiect și a clasificărilor de bibliotecă. Este un aspect important în procesul de *data mining*. TEL ia în considerare următoarele tezaure de vedete de subiect:
  - LCSH
  - RAMEAU
  - SWD
  - BISAC
  - 50 k

și următoarele clasificări zecimale:

- LCC (Clasificarea Bibliotecii Congresului)
- DDC (Clasificarea Zecimală Dewey)
- UDC (Clasificarea Zecimală Universală)

Din clasificarea disciplinelor sunt derivate clasele TEL, clase de vedete de subiect. Maparea efectuată manual a înregistrărilor la disciplinele cercetării științifice nu este fezabilă! Rolul schemelor de clasificare existente este de a ușura sarcina mapării și a o face mai fezabilă.

- Abordări:
  - Crowdsourcing
  - Mapare Manual In-House
  - Mapare Semiautomată
  - Tehnici Machine Learning
  - Crearea seturilor de date:
    - Seturi de date pentru training
    - Seturi de date pentru testare

15:30-15:45

ședința plenară de încheiere a lucrărilor acestei reuniuni, în care au fost rostite discursuri de mulțumire și au fost dăruite buchete de flori organizatoarelor.

### **E. Concluzii**

1. La reuniunea comună au participat 107 delegați din cel puțin 26 de țări europene care au reprezentat următoarele instituții, toate partenere în proiectul Bibliotecile Europeana sau în alte proiecte Europeana, cu excepția lui Clemens Neudecker de la Biblioteca Regală a Olandei, care a reprezentat proiectul european IMPACT:

- Biblioteca Parlamentului Ungar (Ungaria)
- Universitatea din Belgrad (Serbia)

- CERL
- Zentralbibliothek Zürich (Elveția)
- The European Library (TEL – Biblioteca Europeană)
- Biblioteca Națională a Slovaciei (Slovacia)
- Biblioteca Națională a Franței (Franța)
- Universitatea din Lund (Suedia)
- K.U. Leuven (Belgia)
- Fundația Europeană a Științelor
- Biblioteca Națională a Albaniei (Albania)
- Biblioteca Națională a Austriei (Austria)
- Biblioteca Națională a Republicii Cehe (Republica Cehă)
- Biblioteca Națională a Germaniei (Germania)
- Biblioteca Națională a Suediei (Suedia)
- Biblioteca Națională a Olandei (Olanda)
- British Library (Marea Britanie)
- Universitatea Lucian Blaga Din Sibiu (România)
- Biblioteca de Stat a Rusiei (Rusia)
- Universitatea din Viena (Austria)
- Biblioteca Academiei Române (România)
- Biblioteca Universității Ghent (Belgia)
- Biblioteca Națională a Elveției (Elveția)
- Instituto Superior Técnico (Portugalia)
- Biblioteca Parlamentului Național al Gruziei (Gruzia)
- Biblioteca Națională a Țării Galilor (Marea Britanie)
- SalterBaxter
- Biblioteca Națională a Rusiei – St. Petersburg (Rusia)
- Biblioteca Națională și Universitară din Bosnia-Herțegovina
- Biblioteca Națională a Poloniei (Polonia)
- Biblioteca din Cipru (Cipru)
- Europeana
- Biblioteca Națională și Universitară a Sloveniei (Slovenia)
- Biblioteca Națională a Estoniei (Estonia)
- Biblioteca Națională a Moldovei (Moldova)
- Universitatea din Tartu (Estonia)
- Universitatea Complutense din Madrid (Spania)
- Digitaal Ergoed Nederland (Olanda)
- Biblioteca Națională a Spaniei (Spania)
- Bayerische Staatsbibliothek (Germania)
- Bibliotecile Bodleiene, Universitatea din Oxford (Marea Britanie)
- Biblioteca Wellcome (Marea Britanie)
- LIBER
- LIBER – Comitetul pentru Marketing și Comunicații

- Universitatea din Uppsala
- Biblioteca Universității din Uppsala (Suedia)
- Biblioteca Academică a Universității din Talin (Estonia)
- Biblioteca Națională și Universitară St. Kliment Ohridski (Bulgaria)
- Biblioteca Națională a Danemarcei (Danemarca)
- TEL

Se poate observa că la această întâlnire au fost prezenți, atât parteneri mai vechi, cât și mai noi, ai Europeanei, precum și unii parteneri potențiali, din partea altor proiecte europene precum IMPACT etc. Reprezentate au fost atât bibliotecile naționale cât și bibliotecile de cercetare.

**2.** Această reuniune vizează pachetele de lucru 3, 4 și 5 din proiectul Bibliotecile Europeana.

**3.** În urma analizei statistice efectuată de echipa TEL asupra chestionarului din pachetul de lucru 3 și a prezentării defalcate a planului de ingerare a datelor, delegația Bibliotecii Academiei Române a constatat că în proiect instituția nu figurează decât cu imagine, nu și cu full-text. Dată fiind lipsa de infrastructură specializată pe ocr-izare și de experiență în această formă complicată de prelucrare a datelor, este recomandabil să se renunțe la ideea de a mai furniza full-text.

**4.** Biblioteca Academiei Române, în calitatea sa de furnizor de conținut către Europeana, va avea raporturi directe doar cu echipa TEL, care se angajează față de toți furnizorii de conținut din cadrul proiectului Bibliotecile Europeana să asigure recoltarea, maparea, indexarea, gruparea/configurarea (clustering), fațetarea (data mining) și îmbogățirea datelor furnizate de aceștia în vederea pregătirii pentru includerea lor în Modelul de Date Europeana și, în final, pentru ingerarea și agregarea lor de Europeana. Biblioteca Academiei Române nu trebuie să facă altceva decât să livreze metadatele și obiectele digitale exact așa cum le are. Formatul UNIMARC în care sunt lucrate înregistrările bibliografice ale Bibliotecii Academiei Române, făcând parte din familia formatelor MARC, este preferat de TEL, față de DublinCore, format în care lucrează, de exemplu, partenerii noștri de la Biblioteca Universității „Lucian Blaga” din Sibiu. O lipsă a noastră, din perspectiva procesului de *data mining*, este lipsa clasificării din structura înregistrărilor noastre bibliografice, dar un plus este traducerea metadatelor în limba engleză.

**5.** Se pot distribui gratuit programe software și instrumente de digitizare între partenerii proiectului Bibliotecile Europeana. Se practică schimbul de experiență și de atu-uri.

**6.** Conform planului de ingestie a datelor prezentat de echipa TEL, Biblioteca Academiei Române va face o singură livrare către TEL, a tuturor celor 3440 de itemi cu care s-a înscris în proiect de la bun început,

abia în ianuarie anul viitor. Vom fi anunțați prin mail cu o lună înainte să ne vină rândul.

7. În urma discuției purtate de doamna Flori Dumitrescu cu domnul José Borbinha, a reieșit că ne-ar trebui un URL persistent, sau un identificator persistent. A rămas ca doamna Flori Dumitrescu să se consulte în acest sens cu partenerii noștri din Budapesta, de la care provine sistemul ALEPH.

8. Livrabilul 4.1 intitulat „Requirements infrastructure and harvester” (Cerințe pentru infrastructură și recoltarea datelor) a circulat deja și a fost comentat via email și înainte de întâlnire. El urmează să fie citit, eventual tradus în limba română, spre a fi lecturat, și discutat în grup de echipa de proiect a proiectului Bibliotecile Europeana. TEL așteaptă un răspuns de la noi cât mai curând posibil. Dacă nu le vom oferi partenerilor noștri externi nici un fel de *feedback*, acest document va fi adoptat într-o formă care s-ar putea mai târziu să nu ne avantajeze.

9. Dacă proiectul ARROW va fi încununat de succes, atunci ni se va oferi posibilitatea rezolvării automate a problemelor legate de *copyright*.

#### **F. Devizul costurilor deplasării la Haga:**

Participant	Transport internațional: KLM Buc.-Amst. Dus-Întors	Transport interurban: a) tren b) tramvai	Cazare Hotel Sebel: a) cameră Single b) mic dejun	Cina festivă	Diurnă	Total EURO
Flori Dumitrescu	285	Tren: 18, 60 Tramvai: 10	271,62	35	140	696,62
Ligia Caranfil	285	Tren: 18,60 Tramvai 10	257,62	35	140	682,62
Total costuri 4 zile	570	57,20	523,24	70	280	1.373,24

Redactat de Ligia Caranfil, secretară de proiect, la data de 11 iunie 2011.

Data:

Semnătura:

Flori Dumitrescu

Ligia Caranfil