

Parallel Session 1

IMPACT as a Centre of Competence: Towards a Sustainable Centre of Competence in Mass Digitisation of Historical Printed Text

Aly Conteh – BL, and Hildelies Balk – KB

Tuesday 7 April 2009, Aula, 15.30 – 17.00

No.	Name	Organisation	Country	Email address
1	Xavier Agenjo	Fundación Ignacio Larramendi	Spain	agenjo@ono.com
2	Niall Anderson	The British Library	United Kingdom	niall.anderson@bl.uk
3	Hildelies Balk	National Library of the Netherlands	The Netherlands	hildelies.balk@kb.nl
4	Jeannine Beeken	INL - Institute for Dutch Lexicology	The Netherlands	Jeannine.Beeken@inl.nl
5	Fedor Bochow	Bavarian State Library	Germany	Fedor.Bochow@bsb-muenchen.de
6	Rafael Carrasco	Biblioteca Virtual Miguel de Cervantes	Spain	carrasco@dlsi.ua.es
7	Naveen Chamoli	Planman Technologies	India	naveen.chamoli@planmantechologies.com
8	Aly Conteh	The British Library	United Kingdom	aly.conteh@bl.uk
9	Michael Day	UKOLN, University of Bath	United Kingdom	lismd@ukoln.ac.uk
10	Taomir Ebersold	ARPA Data GmbH	Switzerland	info@arpa.ch
11	Berry Feith	National Library of the Netherlands	The Netherlands	berry.feith@kb.nl
12	Neil Fitzgerald	The British Library	Greece	neil.fitzgerald@bl.uk
13	Francisca Hernandez	DIGIBIS S.L.	Spain	francisca.hernandez@digibis.com
14	Chris Honsinger	Certifi Media Inc.	Canada	cwh@certifi-media.com
15	Ádám Horváth	Hungarian National Library	Hungary	adam@oszk.hu
16	Tiina Ison	National Library of Finland - The National Digitisation Centre	Finland	tiina.ison@helsinki.fi
17	Jukka Kervinen	National Library of Finland - The National Digitisation Centre	Finland	jukka.kervinen@helsinki.fi
18	Evelien Ket	National Library of the Netherlands	The Netherlands	evelien.ket@kb.nl
19	Daniel Knippers	LexisNexis Univentio	The Netherlands	daniel.knippers@lexisnexis.com
20	Joachim Korb	Austrian National Library	Austria	joachim.korb@onb.ac.at
21	Marian Lefferts	CERL	The Netherlands	marian.lefferts@cerl.org
22	Agnieszka Leszyska	National Library of Poland	Poland	a.leszyska@bn.org.pl
23	Josep Lladós Canet	Computer Vision Centre	Spain	josep@cvc.uab.es
24	Christian Mahnke	Göttingen State and University Library	Germany	mahnke@sub.uni-goettingen.de
25	Hamid Mehrabi	The Royal Library of Denmark	Denmark	hmr@kb.dk
26	Barbara Muñoz de Solano y Palacios	Spanish National Library	Spain	barbara.munoz@bne.es
27	David Nelson	European Patent Office	The Netherlands	dnelson@epo.org
28	Mr. Roberg	ARPA Data GmbH	Switzerland	info@arpa.ch
29	Vishal Salgotra	Planman Technologies	India	vishal.salgotra@planmantechologies.com
30	David Schmidt	Certifi Media Inc.	USA	davideschmidt@mac.com
31	Henning Scholz	Museum für Naturkunde	Germany	henning.scholz@mfn-berlin.de
32	Klaus Schulz	CIS - University of Munich	Germany	schulz@cis.uni-muenchen.de
33	Simon Tanner	King's College	United Kingdom	simon.tanner@kcl.ac.uk
34	Rubrecht Zaat	National Library of the Netherlands	The Netherlands	rubrecht.zaam@kb.nl
35	Frederick Zarndt	Planman Consulting In	USA	frederick.zarndt@planmanconsulting.com

Document history

Revisions

Version	Status	Author	Date	Changes
0.1	Draft	Niall Anderson	16 April 2009	Created
1.0	Final	Lieke Ploeger	26 May 2009	PDF

Distribution

Version	Date of sending	Name	Role in project
0.1	24 April 2009	Aly Conteh, Michael Day, Joachim Korb, Jeanna Nikolov-Ramírez	Ad hoc conference group
1.0	26 May 2009	IMPACT website	All interested parties

Notes

Introduction: Hildelies Balk (HB)

Moderator: Aly Conteh (AC)

Introduction:

HB restates the idea of IMPACT as a Centre of Competence; mentions comments from conference attendees about European focus of IMPACT; explains that emphasis on Europe is due to funding conditions rather than on specific qualities of mass digitisation in Europe.

Phase 1. (2008-2009) makeup of IMPACT Consortium; mentions contact with other Centres of Competence, including National Digital Information Infrastructure and Preservation Program¹ at Library of Congress.

Phase 2. (2009-10): extension of consortium will feature 3 to 6 partners possibly from East and Southern Europe, which will further both mass digitisation competence through Europe and act as testing ground for the language-independent IMPACT tools.

Phase 3. (2010-11): need to identify business model for self-sustainability. Discusses existing platforms for dissemination / outreach; existing relationships between IMPACT consortium and partners in MLA community, R&D community, and service providers.

Main session:

AC mentions difficulty of building roadmap for sustainability from ground up. Explains that BL team started process by analysing project outputs that could be leveraged into sustainable growth of IMPACT, and continued by identifying stakeholders (potential and actual) of IMPACT. Refers to list of potential stakeholders and asks if audience considers themselves part of any one – or none.

¹ <http://www.digitalpreservation.gov/>

Klaus Schulz suggests "consumers" not an adequate description of correctors, enhancers, online community, etc. On same point, Rose Holley [RH] explains that the Australian Newspaper Project has used its online community for testing and enhancement of service, and could be used to steer Wikipedia-style funding drives.

Rafael Carasco (Universidad Europea Miguel de Cervantes, Alicante) suggests that term "software providers" can be segmented into commercial providers and open source community.

Simon Tanner [ST] suggests addition of publishers and search engine operators to list. HB mentions that search engine providers have contacted IMPACT about possibility of testing OCR tools.

David Nelson [DN] from European Patent Office feels that bodies such as his and UNESCO are not included in initial list of stakeholders.

In context of IPR, Jeannine Beeken [JB] of INL suggests legal profession as new stakeholder.

Tiina Ison [TI] (National Library of Finland) suggests metadata standards organisations as stakeholders; also recommends thinking of end users ("consumers") as Communities of Practice.

Richard Helle [RHE] of CCS suggests addition of term referring to issues of long-term storage and preservation. TI agrees. HB mentions Planets² as Centre of Competence for preservation rather than IMPACT. RHE talks about competing technical standards in MLA community, and the difficulty service providers have in satisfying specifications as a result.

RHE: where does Google fit in list of stakeholders? Mentions that Google are service providers and digitisation partners, but also digital publishers in their own right, and developers of open source software.

Capability, Research, Tools and Partnership

AC introduces concepts of Capability, Research, Tools and Partnership.

Capability: deliverables from project that capture expertise within consortium and can be used to drive sustainability.

Research: emphasis on advocacy more than funding.

Tools: active management of toolset, including later iterations and recognition of obsolescence. AC asks what's missing from model.

Frederick Zarndt [FZ] (Planman): project/programme management missing from list; what about communications tools for distributed projects? AC asks if FZ is talking about developing tools for managing workflows. FZ says yes.

² <http://www.planets-project.eu/>

JB concerned about how users will interact with Centre of Competence: how integrated will Centre of Competence be if so many people and institutions are involved? Suggests single portal for delivery of digitised material throughout the consortium. AC suggests access to material falls under Europeana, while IMPACT hopes to make the growth of Europeana content easier. HB suggests that consortium members should have identical Centre of Competence pages on their website, driving traffic to each other. Daniel Knippers (Lexis Nexis) suggests Google as ready made indexer of Centre of Competence content.

Marian Lefferts [ML] of CERL: right material in right environment with right standards will ensure findability, and enable intelligent use of tools by research community.

RH suggests that IMPACT should explicitly foster innovation, under either Capability or Research headings.

AC: question to the floor – do we need a centre of competence? ST notes historic danger of networks of excellence talking only to themselves. Need to ensure engagement outside consortium.

AC: question to the floor – what are delegate expectations of IMPACT Centre of Competence? RE: Centre of Competence as creator and guardians of exemplary guidelines about technical specifications for mass digitisation. FZ: Centre of Competence as driver of financial efficiency, giving model specifications that drive costs down for both service providers and MLAs. Need for greater technical standardisation echoed by several delegates.

RHE affirms applicability of IMPACT tools outside Europe; asks how would services be accessed (e.g. dataset)?

HB: if IMPACT expertise were used to further research outside EU, this may bring IMPACT into conflict with the terms of its funding, though to her knowledge the EC had never placed any such strictures on its projects.

Neil Fitzgerald [NF]: service providers needed for evidence-based research, so service providers would have to be in on ground floor.

DN asks about attitude of EC to Centre of Competence idea. HB explains that there is an expectation that IMPACT should self-sustain beyond 2011, but that this is not an explicit project deliverable. Mentions recent EC review, where IMPACT was urged to take Centre of Competence idea further: suggesting strong support within EC.

ST: content should speak for itself: pressure on big institutions to put large volumes of material online, but not necessarily the best content – what of smaller countries/collections?

TI: suggests process mapping approach to management in libraries (as recently done at Library of Finland); process mapping a move away from traditional hierarchal understanding of library management – concentrating on actual dependencies rather than organisational boundaries.

Taomir Ebersold: agrees with what NF said. In many projects there is a lack of realism and understanding between technical and library partners. David Schmidt (Certifi): re-emphasises that uniform language is needed for project management and technical standards.

Jukka Kervinen (Library of Finland): suggests that IMPACT could engage in equipment "reviewing". NF notes danger of building expectations too high.

Sustainability:

AC asks how IMPACT might be sustained beyond 2011.

RE says funding may be available from consortium members and commercial partners. ML suggests IMPACT could become a member of Liber³. JB recounts her experience with Liber as part of CLARIN⁴ project: CLARIN stakeholders resisted becoming a Centre of Competence immediately, but persuaded national governments to fund consortium towards complete self-sufficiency. RH suggests that after 2011 possibility of extra European partnerships; may be necessary to charge for use thereafter. ST calls this a Community Trading model. FZ suggests user groups raising money as in Australian Newspaper project; acknowledges this as a bit of a stretch. DS suggests that rather than looking just at funding, consortium members should ask what they individually wish to achieve within IMPACT: results orientation to drive funding. Library of Finland suggests bidding for funding with other Centres of Competence.

AC thanks all. Roadmap slated for third quarter 2009, asks permission of audience to seek further comments offline; page for feedback on IMPACT website forthcoming.

³ <http://www.libereurope.eu/>

⁴ <http://www.clarin.eu>