# the well-behaved document

Presented by John W. Miescher Bizgraphic – Geneva On Thursday, October 06, 2011

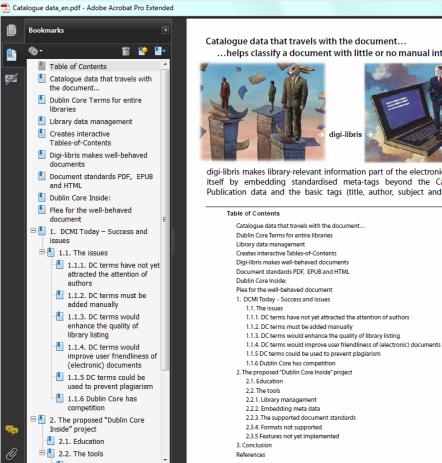
# the well-behaved document

is an electronic document that is both user friendly and library friendly

#### is easy to read and to navigate

#### it has bookmarks

and an interactive table-of-contents



...helps classify a document with little or no manual intervention digi-libris makes library-relevant information part of the electronic document itself by embedding standardised meta-tags beyond the Catalogue-in-Publication data and the basic tags (title, author, subject and keywords). 1.1.1. DC terms have not yet attracted the attention of authors 1.1.3. DC terms would enhance the quality of library listing

- O X

is practical to consult and arouses more interest

4

# the well-behaved document

is an electronic document that is both user friendly and library friendly

it includes useful metadata as part of the document itself

to help classify it with little or no manual intervention

for personal data mining and re-use

| 🔁 Cata | logue o | data_en.j                | odf - Adobe Acrobat P            | Pro Extended                          |                 |   |  |                       |   |
|--------|---------|--------------------------|----------------------------------|---------------------------------------|-----------------|---|--|-----------------------|---|
|        | Bookr   | Table                    | f Contante<br>Current Item = CP3 |                                       |                 | data that travels with<br>classify a document v | the document<br>vith little or no manua        | l intervention        | E   |
| 1      |         | Catalc<br>the do         |                                  | n Preferences                         | Dublin Core T   | A 7   |  |                       | <b>*</b>  |
|        |         | Dublin                   | Object/PDF Prop                  |                                       | ore Elements    |   | Local/private content                          | Imment                |   |
|        |         | librari                  | objection                        | Dubin C                               | ore clements    | Determs + Qualyners                             | Local/private content                          | Import                |   |
|        | -       | Librar<br>Creat          | Item                             | CP152_sp.pdf                          |                 |   | 06 Informe anual del ONUSDA<br>Ren valv d dans | File Crea Date        | 2007-07-05  |
|        |         | Table<br>Digi-li         | Location                         | E:\BizgraLib\Test                     | Dir\N 2\N 2a\   |   |  | File Modif Date       | 2010-05-15  |
|        |         | docun<br>Docur<br>and H  | Object                           | PDFb                                  |                 | <b>_</b>  | in a   | Number of Pages       | 72  |
|        |         | Dublir<br>Plea f         | Image<br>(local)                 |                                       |                 |   |  | File Size             | 1821722   |
|        |         | docun<br>1. DC<br>issues |                                  | Namespace = htt                       |                 | 1.        | DF   |                       |   |
| -      |         | 1.1                      | Title                            | Informe anual de                      | ONUSIDA 200     | 6 Hacer rendir el dinero                        |  |                       |   |
|        |         |                          | Creator                          | ONUSIDA 1                             |                 |   |  |                       |   |
|        |         | -                        | Description                      | Creado en 1995 p                      | or medio de ur  | na resolución del Consejo                       | Económico y Social de la                       | as Naciones Unidas,   | , y lanzado en enero de 1996,                               |
|        |         | -                        |                                  | el Programa Con                       | junto de las Na | ciones Unidas sobre el VI                       | H/SIDA (ONUSIDA) es una                        | a iniciativa conjunta | a innovadora de las Naciones<br>del sistema de las Naciones |
|        |         |                          |                                  | Unidas que conce<br>Unidas para respo |                 | zos y recursos de la secre                      | taria del UNUSIDA y de di                      | ez organizaciones o   | del sistema de las inaciones                                |
|        |         | -                        |                                  |                                       |                 |   |  |                       |   |
|        |         |                          |                                  |                                       |                 |   |  |                       |   |
|        |         |                          | Keywords                         | PCB<br>cosponsors                     |                 |   |  |                       |   |
|        |         |                          |                                  | annual                                |                 |   |  |                       |   |
|        |         | -1                       |                                  | financials<br>three ones              |                 |   |  |                       |   |
|        | 8       | 2. The                   |                                  | strategic framew                      | ork             |   |  |                       |   |
| ۶Þ     | _       | Inside<br>2.1            | Web Copyright                    |                                       |                 |   |  |                       |   |
| Ø      | -       | 2.2                      |                                  |                                       |                 | reset to last s                                 | aved New                                       | Update file           | Update list <b>close</b>                                    |

are more attractive for readers to keep and refer to have a better chance of being found (in search engines)

### Who needs it?

Information creator

author, researcher

Information agent

Information consumer

> faculty, publisher, self, librarian, archivist, digitizer, asset manager, etc.

you, me, students, scholars, non-scholars (min. 75%)

### about Open Access

Open Access is migrating from a buzzword to a quasi standard for scientific publishing.

While emerging concepts like Nanopublikationen, Semantic Publishing, Open Data, Enhanced Publications and Research Objects all have merits in scientific context, Google Scholar might end up stealing the show, even in research.

Great efforts are being made to streamline ontologies (SPAR, CiTO), semantics (CERIF, CRIS), protocols and data integration (SWORD, SHERPA/RoMEO), university/research libraries grouping together to standardize and coordinate their repositories and opening data licensing (LIBER, CENL, CERL, OAPEN, COAR, openAIRE, openDOAR etc.) and

metadata harvesting and interoperability becoming an increasingly important issue (OAI-PMH, MARC, Dublin Core).

All this is undertaken by university and library professionals for the benefit of university and library professionals, it seems.

While the benefits to the individual information consumer appear to be obvious, we were not able to find many pieces of instruction targeted at the end user on how to best profit from these efforts and what skills he/she needs to develop in order to maximize his/her search for information.

### why does he/she need it?

The Information consumer...

looks-up data or collects information

author, researcher, scholar or occasional writer.

searches the Web finds hundreds of references

spends time collecting, sorting and referencing citations and data from mostly unstructured text

must decide on relevant and essential information

and would find embedded metadata most helpful

### if you are doing research work

### if all documents were "well-behaved"...

had bookmarks and interactive tables of contents and had meaningful embedded metadata

### the consumer

#### would have easy access to information

to refer to and to consult off-line sources He/she has acquired or downloaded

### and the writer of a thesis, a report or a paper

wouldn't waste time in search of relevant information, collecting citations and assembling data

### about embedded metadata

The average information consumer...

is primarily interested in the descriptive metadata and less in the structured and administrative metadata

does not care about semantics, namespaces and refinements

the Dublin Core terms (DCMI interoperable online metadata) is probably the best option for embedding useful metadata

the format is widely recognized and well documented

can easily be embedded in PDF, EPUB and HTML files

in PDF files it is already included as standard

### more about Dublin Core

DC terms have not yet attracted the attention of authors authors are non-librarians authoring tools allow at best only basic metadata the discipline to fill in even these basics is not there 15 CD-elements and 55 DC-terms can lead to confusion

#### Dublin Core is not alone

there are lots of standards issued by major libraries, universities, school authorities and other interest groups MARC 21 is used by major libraries for their own repositories Unfortunately it is not very compatible with the Dublin Core standard and the *MARC to Dublin Core Crosswalk* has its limits

# embedding Dublin Core tags

#### nestor-handbuch\_benutzerfreundlicher.pdf

| mad | ×            |  |
|-----|--------------|--|
|     |              |  |
|     | $\mathbf{b}$ |  |

| PDF                  | Description<br>digi-libris properties | Advanced  | ۲     |
|----------------------|---------------------------------------|---|-------|
|                      | Adobe Stock Photos                    | DDE Branatias (adf. http://as.adaba.com/adf/1.2/)   |       |
|                      | IPTC Contact                          | PDF Properties (pdf, http://ns.adobe.com/pdf/1.3/)     XMP Core Properties (xmp, http://ns.adobe.com/xap/1.0/)    |       |
|                      | IPTC Image                            | WMP Core Properties (xmp, nccp.//ns.adobe.com/xap/1.0/)   |       |
|                      | IPTC Content                          | Dublin Core Properties (dc, http://purl.org/dc/elements/1.1/)   |       |
|                      | IPTC Status<br>Camera Data 1          | Core Properties (dc, ncc)./pdn.org/dc/elements/1.1/)  |       |
|                      | Camera Data 2                         |   |       |
|                      | Categories                            |   |       |
|                      | History                               | dc:title (alt container)  |       |
|                      | DICOM<br>Audio Data 1                 | dc:dde (al: container)  |       |
|                      | Audio Data 1<br>Audio Data 2          | dc:publisher (bag container)  |       |
|                      | Video Data 1                          | dc:creator (seg container)  |       |
|                      | Video Data 2                          | dc:subject (bag container)  |       |
|                      | Advanced                              | □ http://purl.org/dc/terms/   | 1     |
|                      |                                       | dcterms:available: 02.08.2010   | 0.40  |
|                      |                                       | dcterms:created: 19.07.2010   | mer   |
|                      |                                       | dcterms:conformsto: well-behaved document description (see http://www.bizgraphic.ch/digi-libris/digi-libris.html) | -cap. |
| most scal            |                                       | dcterms:language (bag container)  |       |
|                      |                                       | dcterms:rights (alt container)  |       |
|                      |                                       | dcterms:audience (seq container)  |       |
| A arabat ®           |                                       | http://ns.adobe.com/pdfx/1.3/   |       |
| Acrobat <sup>®</sup> |                                       | +ttp://adsml.org/xmlns/   |       |
|                      |                                       | ttp://gwg.org/jt/xmlns/   |       |
|                      |                                       | + http://ns.adobe.com/illustrator/1.0/  |       |
| has a spa            |                                       | 🗄 http://digi-libris.org/xmlns/   |       |
|                      |                                       | Adobe Photoshop Properties (photoshop, http://ns.adobe.com/photoshop/1.0/)  |       |
|                      |                                       | ian http://ns.adobe.com/png/1.0/  |       |
| in alunda a i        |                                       | TIFF Properties (tiff, http://ns.adobe.com/tiff/1.0/)   |       |
| includes i           | Powered By                            |   |       |
|                      | XMD                                   | Replace Append Save Delete  | e     |
|                      |                                       |   |       |
| but no too           |                                       | OK Cance  | el    |
|                      |                                       |   |       |
|                      |                                       |   |       |

### embedding Dublin Core tags

#### EPUB

the new standard for ebooks. is not proprietary some layout programs (e.g. InDesign) can generate it directly Handled by most major ebook readers *Most examples we have seen use dc:terms correctly.* 

#### HTML

the most uncontrolled format of all there are hundreds of applications that can create it, *each with his own flavor of syntax and scripts* 

By far not all originators use <dc...> or <dc.terms...> for metadata and often those that do add fancy designations of their own.

Can be virtual or created on the fly, e.g. in response to a query

### helpful software tools

### digi-libris to organize your knowledge base and collections of documents

|                  | digi-libris Library Manager   |  |  |  |  |  |  |
|------------------|---|--|--|--|--|--|--|
|                  | File Collection Edit Settings Action View Help  |  |  |  |  |  |  |
| add physical or  | Collection: anton Automation AR9.pdf all readable   |  |  |  |  |  |  |
|                  |   |  |  |  |  |  |  |
| electronic items | abdur.pdf Object/PDF Properties Dublin Core Elements DCterms + Qualyfiers Local/privat  | Laurente Laurent   |  |  |  |  |  |
| o your           |   |  |  |  |  |  |  |
|                  | Amphora wi Item  CP152_sp.pdf   | File Crea Date 2007-07-05                                    |  |  |  |  |  |
| collection       | Angels & de     Location E:\BizgraLib\TestDir\N 2\N 2a\   | File Modif Date 2010-05-15                                   |  |  |  |  |  |
| and see the      | Automation Object PDFb  Automation  | Number of Pages 72   |  |  |  |  |  |
| embedded         | CP150xy_ru. Image<br>CP152_sp.pc (local)  | File Size 1821722  |  |  |  |  |  |
| metadata         | Image: Second system     Image: Second system       Image: Second system     Second system       Image: Second system     Namespace = http://ns.adobe.com/pdf/  | fault  |  |  |  |  |  |
| Prove I          | Title Informe anual del ONUSIDA 2006 Hacer rendir el dinero   |  |  |  |  |  |  |
| or edit and      | Dublin Core   |  |  |  |  |  |  |
| re-embed the     | ExecSumm_   | fault  |  |  |  |  |  |
|                  | Greville-Pere     Description     Creado en 1995 por medio de una resolución del Consejo Económico     el Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA (ONU   | y Social de las Naciones Unidas, y lanzado en enero de 1996, |  |  |  |  |  |
| metadata         | Unidas que concentra los esfuerzos y recursos de la secretaría del ONU  |  |  |  |  |  |  |
|                  | 🔲 le documen Unidas para responder al SIDA.   |  |  |  |  |  |  |
| o make a         | LofC Test     Image: A constraint of the second secon |  |  |  |  |  |  |
| document         | ✓ nestor-hand<br>✓ nestor-hand<br>Keywords PCB  |  |  |  |  |  |  |
|                  | O37.pdf cosponsors  |  |  |  |  |  |  |
| ibrary friendly  | Plaidoyer pc     annual     Plea for the     financials   |  |  |  |  |  |  |
|                  | Proverbes_si three ones   |  |  |  |  |  |  |
|                  |   |  |  |  |  |  |  |
|                  | Web Copyright MetaData of E:\   |  |  |  |  |  |  |
|                  | reset to last saved   | New Update file Update list close                            |  |  |  |  |  |
|                  |   |  |  |  |  |  |  |

download a trial version of this software from http://digi-libris.com

# post production software

### to enhance existing PDF files, e.g. third party or digitized ones you

#### can...

copy/paste or type a table of contents

then automatically generate and edit a bookmark list

and send \* bookmarks and links to a PDF to to build an interactive table of contents

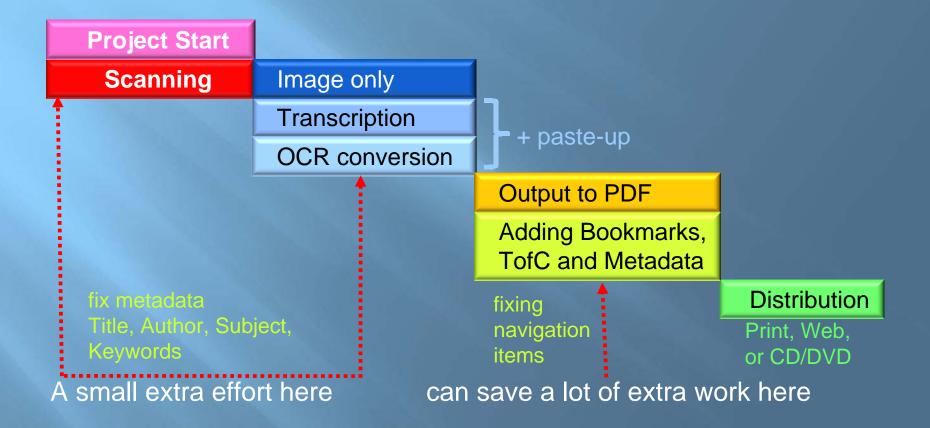
to make a document user friendly

|                         | erzeichnis in PDF erstellen  |
|-------------------------|--|
|                         | führen Lesezeichen-Liste   |
| *.pdf E:\BizgraLib\Test |  |
| Seite m.Inhaltsv        | Lesezeichen und Inhaltsverzeichnis in PDF erstellen  |
| 1 h h                   | Datei Bearbeiten Ausführen Lesezeichen-Liste   |
| Inhaltsverzeichnis List | *.pdf E:\BizgraLib\TestDir\N1\CP152_en.pdf PDF öffnen suchen nach PDF nächste  |
| Table of Contents       | Seite m.Inhaltsverz. 5 📮 Delta (angezeigte ± phys.Seiten-Num.) 2 🚔 Anz.Spalten 1 🚔 Randabst.  -> 54 48 <-  |
| Foreword 5              | Inhaltsverzeichnis Liste Lesezeichen Stufe+ Stufe- Pagin.+ Pagin. PDF beschicken 💿 Lesezeich.+Links 🔘 nur Lesezeichen 🔘 nur Links 🔘 blinde Lin   |
| Overview of UNAIDS      |  |
| AIDS epidemic in 200    | level target caption   |
| Chapter One: Leaders    | 0 -5 Table of Contents<br>0 7 Foreword   |
| High Level M            | 0 9 Overview of UNAIDS   |
| Strengthening           | 0 11 AIDS epidemic in 2006   |
| Women and               | 0         13         Chapter One: Leadership and advocacy           1         13         High Level Meeting on AIDS  |
| Youth and ch            | 1         13         High Level Meeting on AIDS           1         13         Strengthening the UN response to AIDS   |
| Security and            | 1 14 Women and girls   |
| Chapter Two: Strateg    | 1 16 Youth and children  |
| Universal acc           | 1         19         Security and the humanitarian response           0         21         Chapter Two: Strategic information and policies   |
| The "Three O            | Chapter i wor strategic information and poinces     21 University of counter of counter of the counter of |
| Intensifying H          | 1 24 The "Three Ones"  |
| Increasing ac           | 1 25 Intensifying HIV prevention   |
| Human rights            | 1         27         Increasing access to HIV treatment           1         30         Human rights, gender and law  |
| Mainstreamir            | 1 32 Mainstreaming AIDS in development processes   |
| Chapter Three: Survei   | 0 35 Chapter Three: Surveillance, monitoring and evaluation  |
| The epidemiology of     | 0 35 The epidemiology of HIV   |
| Supporting country-le   | 0         37         Supporting country-level monitoring and evaluation systems           0         38         Simplifying and harmonizing indicators for monitoring programmes  |
| Simplifying and harm    | 0 -5 Monitoring progress on the 2001 Declaration of Commitment on  |
| Monitoring progress     | 0 39 HIV/AIDS and the 2006 Political Declaration on HIV/AIDS   |
| HIV/AIDS and the 200    | 0         43         Chapter Four: Civil society engagement and partnerships           0         43         Working with civil society   |
| Chapter Four: Civil so  | 0 44 People living with HIV  |
| Working with civil soc  | 0 48 Working with faith-based organizations  |
| People living with HIV  |  |
| Working with faith-ba   | 0         51         Chapter Five: Resource mobilization           0         51         Estimating resource needs and tracking expenditures  |
| Partnerships with labo  | 0 54 Increasing access to funding  |
| Chapter Five: Resourc   | 0 55 Supporting implementation   |

download a trial version of this software from http://digi-libris.com

### good planning saves time and money

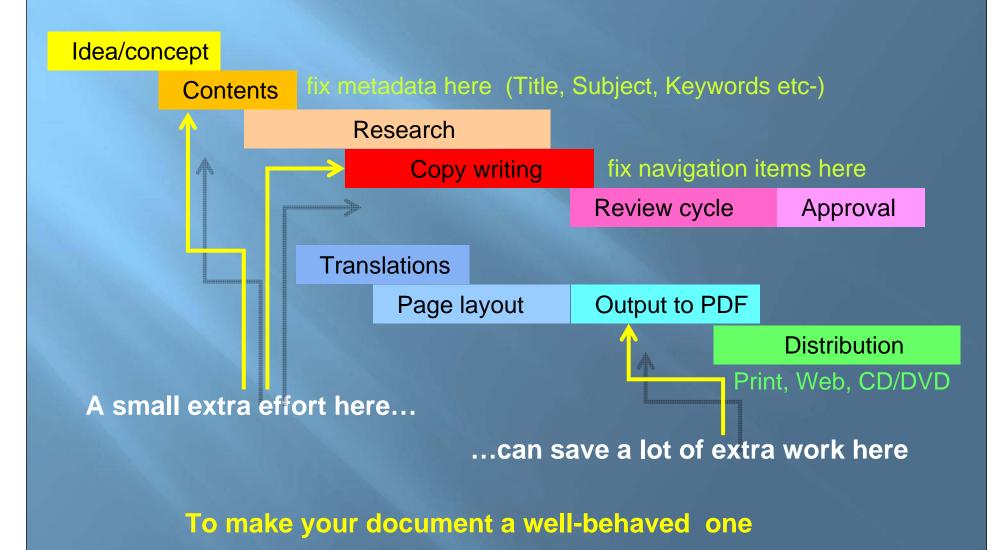
Consider the workflow of a typical digitizing project



to make your document a well-behaved one

### good planning saves time and money

Consider the workflow of a typical document creation process



### Well-behaved documents...

cater to the needs of (and empower) the information consumer have a better chance of being found (in search engines) facilitate personal data mining and re-use are practical to consult and arouse more interest are more attractive for readers to keep and refer to repeatedly are a step in the direction of the semantic web allow processing of library records with little/no manual intervention can save time and money if planned and implemented accordingly

# the well-behaved document has Dublin Core inside

Introducing "Dublin Core inside" as a new standard and mark of quality for truly well-behaved documents,

to be recognized worldwide by information providers and information consumers alike.

#### How?

by using our influence, relations and know-how with authors, publishers, aggregators and content providers

by helping all stake holders to understand and appreciate the added value of well-behave documents and to prepare all their documents accordingly.