



Internationaler Archivkongress 2004

23.-29. August - Wien - Österreich

**Archive, Gedächtnis und Wissen**



# Describing Archives Online: Tools for Authoring Online Finding Aids

Anne van Camp

Hubert Wajs

Werner Engel

Angelika Menne-Haritz

## The Polish Road to EAD

Hubert Wajs

Central Archives of Historical Records, Warsaw, Poland

Archival community – as far as I know – looks at standards and automation with some kind of hesitation. Such situation was very good described by Professor Peter Horsman on European Union Archive Network (EUAN) web site<sup>1</sup>. He hoped that old motto: “my archives are unique” is going to be the past. I remember my own experience with the first Polish translation of General International Standard Archival Description – ISAD(G) in 1995. Translation of the second version of ISAD(G), which had just been finished, was done in a different climate. Similar attitude of hesitation could be noticed towards Encoded Archival Description (EAD). My own personal reaction, when for the first time I saw the template in SGML, was borrowed from the film for children >>Bob the Builder<<: “Can we fix it?” In this film the answer is: “Yes, we can!” and it is a real answer – “we can”. First of all I would like to express my thanks and acknowledge my debt to many colleges archivists, who helped me on my road to EAD. I wonder, or even doubt, if one could do it alone. With the goodwill and help of the President of Polish State Archives I had a chance to exchange opinions on EAD with Swedish, German, Dutch and English archivists. During my four days studying visit in Archieschool in Amsterdam Professor Peter Horsman told me to start encoding, even on paper: “Just do it, You must feel it.” Perfect and profound advice. These four days helped me a lot. You can not understand this structure without doing it by your own hands, putting tags in proper order and filling in data between them into the template. Professor Daniel Pitti, whom I never met so far, patiently answered on my successive questions and problems with e-mails of suggestions, corrections and advices. William Stocking suggested to translate “RLG best practice guidelines for Encoded Archival Description“, which I have done<sup>2</sup>.

In our “perfect” archival world, in repositories, we keep not only records (as our holdings), but finding aids as well, to know what these records were about. Nowadays these finding aids (inventories, registers. etc.) and of course archival holdings, should be accessible for all. Inventories should be presented in electronic format on the web, to guarantee easy access to records. Descriptive information is a key to the holdings. Without inventories access, which is right now one of the most important issue for archives, would be not possible<sup>3</sup>. “EAD is the first data structure standard to facilitate distribution via Internet of detailed information about archival collections and fonds” – as one could read in the preface to the *Application guidelines to EAD* (1999). This is what EAD is for. In my, Polish “perfect” archival world, only the small fraction of finding aids is till now in electronic format. I hope – and in this point I am rather optimistic – it is around 15% of all our inventories. The rest have to be put into electronic format – we call it retrospective conversion of existing finding aids. And once again: EAD is just for such activities.

Speaking of standards in electronic environment, sooner or later, will mean - metadata. The term had been familiar in librarians’ world, but in the first half of the last decade of 20<sup>th</sup> century it was rather something new for archivists. In 2000 we hold in Poland a conference on metadata; around 50 archivists, who interested in electronic records, descriptive standards etc., discussed, what was done by the others – EUAN, EAD, CEDARS etc. The next symposium, a month later was dedicated to EAD – Professor Peter Horsman told us about his ideas (pros and contras) in “EAD – Magic word or terror?” Spring on past year we had two conferences (Polish and international) on EAD, in the framework of “Delos” project. The material was published both in paper and in electronic formats on the website<sup>4</sup>. This was a theoretical base. I think some theory and conceptual work were needed for a good start. Introduction of ISAD(G) in 1995 as a schema for description in Central Archives of Historical Records (my workplace) was also very useful and helpful.

Crown Chancery Public Register (*Metrica Regni*) – very important historical fonds, with typical structure (series, files [books], and items) and detailed modern inventory<sup>5</sup>, was chosen for the first, pilot project. The rights

<sup>1</sup> Peter Horsman, *Metadata and Archival Description*, [http://www.euan.org/euan\\_meta.html](http://www.euan.org/euan_meta.html).

<sup>2</sup> Wskazówki dotyczące „dobrych praktyk” przy wdrażaniu EAD opracowane przez grupę doradcą RLG, s. 77-100, in: *Archiwa w postaci cyfrowej. Standaryzacja. Od ISAD(G) do EAD. Materiały międzynarodowych warsztatów Delos CEE* [Proceedings of the International Delos CEE Workshop „Standardization. From ISAD(G) to EAD. Implementation and Best Practice”], pod. red. E. Rosowskiej, Warszawa 2003.

<sup>3</sup> Angelika Menne-Haritz, *Access – the reformulation of an archival paradigm*, *Archival Science*, no. 1, 2001, pp. 57–82; Polish translation in *Archicon* no 104, 2002, pp 68-95.

<sup>4</sup> On the web site of Central Archives of Historical Records – <http://www.archiwa.gov.pl/agad/ead/ead.html>; see also footnote 2 - Proceedings of the International Delos CEE Workshop.

<sup>5</sup> *Inwentarz Metryki Koronnej. Księgi wpisów i dekretów polskiej kancelarii królewskiej z lat 1447 – 1795*, Opracowały Irena Sułkowska-Kurasiowa i Maria Woźniakowa, Warszawa 1973.

and privileges granted by Polish kings' were recorded – probably from the end of 14<sup>th</sup> century – in the books of Crown Chancery Public Register. The oldest, existing till now books, from the series of registrations (*Libri Inscriptionum*), began in the middle of 15<sup>th</sup> century; the series was continued till the fall of Polish Republic at the very end of 18<sup>th</sup> century. In these books were written down copies of all important kings' charters of permanent value like: donation of estates and landed properties, nominations for the offices, military ranks, etc. There are also highly detailed registers for each record – abstract or brief notes, describing what each and singular record refers – to the earliest part of that series: from second half of 15<sup>th</sup> till 70s. of 16<sup>th</sup> century. These registers or “summaria” – *Matricularum Regni Poloniae Summaria* – were published during last 100 years<sup>6</sup>. Another series, gathering diplomatic records – from the beginning of 16<sup>th</sup> century – is called: legations (*Libri Legationum*). Then in 17<sup>th</sup> century Chancery began series for registration of records, which were sealed in Crown Chancery (*Sigillata*) and in 18<sup>th</sup> century other books for Chancellery records issued by Chancellors. Within the fonds there were also books containing records of kings' courts of justice; however, in most of them part courts' records were destroyed during the 2<sup>nd</sup> WW. The language of most of the books (and records) – might be, except 18<sup>th</sup> century – was Latin, but there were also records in Polish, German and Old Ruthenian. There are 790 books in all series, in registrations – 361 books. For the first half of 16<sup>th</sup> century there are about 30 thousands records in 90 books.

Because of its' great historical value for Polish Kingdom, archival structure and existing finding aids *Metrica Regni* was chosen for the pilot project to introduce EAD for describing finding aids in Central Archives of Historical Records in Warsaw. The project was started in the fall of 2003. The first idea was simply to rewrite the existing inventory partly into MS/Word editor and partly into MS/Access database – at the series and file (book) levels; wrappers: <c01 level = “series”> and <c02 level = “file”>. The general information on the series level was rewritten into MS/Word editor and using past/copy put inside the wrappers: <scopecontent>, <bioghist>, <bibliography> etc. In the same way the introduction to the whole fonds was rewritten and put inside the <archdes> wrapper. The table presents the way of conversion from the database into xml file for the file (book) level.

Table 1. EAD – ISAD – Database for <c03 level="file"> - ISAD 3.1.4.

N <sup>o</sup>	EAD name after conversion	ISAD(G) v2	Type	Polish name of the field
1	<unitid>	ISAD 3.1.1	Numeric	Sygnatura
2	<unitdate>	ISAD 3.1.3	Note	Daty
3	<unittitle>	ISAD 3.1.2	Note	Tytuł
4	<langmaterial>	ISAD 3.4.3	Text 50	Język
5	[formerid] <sup>7</sup>	ISAD 3.1.1	Note	Dawne sygn.
6	<physdesc>	ISAD 3.1.5	Text 250	Opis zewnętrzny
7	<scopecontent>	ISAD 3.3.1	Note	Zawartość/treść
8	<note>	ISAD 3.6.1	Text 250	Uwagi
9	<accessrestrict>	ISAD 3.4.1	Text 250	Udostępnianie
10	<altformavail>	ISAD 3.5.2	Text 50	Mikrofilm

The input of data was partly done in the Central Archives of Historical Records by a staff person (Ms. Katarzyna Janowska-Kucio), partly by a person from labour offices, during her half year practice in the Archives (Ms. Aleksandra Siwek) also in some part by students from one of the high schools in Warsaw (ZNP). The most important part of the task – proofreading, checking of accuracy and correction of errors was done by archivists: Ms. Urszula Kacperczyk, Mr. Michał Kulecki, Ms. Inga Stembrowicz and me. At the beginning of the year 2004 there was a possibility to add highly detailed, lower level – item – abstracts of the single records, as <c03 level = “item”> wrapper, for the oldest books from 15<sup>th</sup> century. These books and records are in Latin. The input of that data was done by a person from labour office (Mr. Witold Poletyło). Central Archives of Historical Records began cooperation with the seminar of Professor Wojciech Krawczuk from Jagiellon University in Krakow, who is actually the leading specialist in the field of kings' chancery of modern times<sup>8</sup>. He also published registers for 3

<sup>6</sup> *Matricularum Regni Poloniae Summaria*, ed. T Wierzbowski and others, vol. I – VI, Warszawa 1905 – 1999.

<sup>7</sup> The name is not from EAD, it is changed on <unitid type="former">

<sup>8</sup> W. Krawczuk, *Metryka Koronna za Zygmunta III Wazy*, Kraków 1995; W. Krawczuk, *Metrykanci koronni. Rozwój registratury centralnej od XVI do XVIII wieku*, Kraków 2002; *Księga wpisów podkanclerzego Tomasza Zamoyskiego z lat 1628 - 1635*, opr. W. Krawczuk, *Sumariusz Metryki Koronnej*. Seria Nowa, t. I, Kraków 1999; *Księga wpisów kanclerza Jana*

books from 17<sup>th</sup> century; his abstracts are in Polish. We hope to broaden this cooperation into collaborative grant project. "Quite easy" – because they are in printed format – we could apply around 42 thousands abstracts. Money is the only need. It is possible to add digitized images of the pages of the books into EAD description. This idea, though very interesting, needs much more money, work (scanning from black and white microfilms), hardware (discs space) and software (browsers). Examples of that possibility – prepared for some books from *Metrica Regni* (MK 16) – one could soon find in the project: Polish Internet Library, where descriptions would be combined with images.

Regarding technical side of the project, the most important factor is web site of Library of Congress dedicated to EAD<sup>9</sup>. One could find there and download basic information needed to work with EAD project. You could download DTD instance for the 2<sup>nd</sup> version (version 2002) of EAD; use examples and tag library to consult your structure (template). Two eminent works of Michel J. Fox are there: "Cookbook" and stylesheets (for the 1<sup>st</sup> version). These stylesheets, with slight emendations, could cooperate with the 2<sup>nd</sup> version. The way of presentation depends on your ideas and could be done in different ways (also: Cascading Style Sheets or HTML). After visiting Library of Congress web site dedicated to EAD all one needs to build a template is XML editor. One of the recommended products is XMetaL; however, for my archival institution it is a bit too expensive. I use Polish HTML editor – "Pajaczek" (Little Spider). It could read XML tags as well.

The next step, having already all these mentioned above: DTD from Library of Congress web site and editor, was input of (structured) data. The data could be extracted from MS/Access database directly into an EAD template, if it is structured according to ISAD(G) (see table 1). The data from MS/Access database were exported into xml file automatically. Some changes were done before export – names of the fields in database were taken from EAD, to have EAD tags after conversion, other in HTML editor (putting <p> after <scopecontent> etc. using Find/Replace function). It will depend on the structure of the fonds, as it is foreseen in EAD, which wrapper will be adopted for which level of description; in this project wrapper <c01> was for series, wrapper <c02> for file (book) and wrapper <c03> for item (abstract). In this project we were happy to have general descriptive information, but in other retrospective conversion activities one could be faced with the lack of them and it will be real problem. Date and structure were corrected and checked during all these steps. When xml file was ready, it needed to be validated against the EAD DTD. I use – thanks to Professor Pitti's suggestion – free ware xmlvalid by Elcel Technology Ltd. It is MS/DOS product. After series of validations and corrections, the last thing to be done was presentation. I used – with small changes – Mr. Fox's stylesheet as a delivery method. On the 29<sup>th</sup> July 2004 the inventory of Crown Chancery Public Register in EAD was visible on the web site of Central Archives of Historical Records<sup>10</sup>. It was xml file with xsl stylesheet; There are no search engines or any other indexing tools so far.

The first inventory in EAD is already done, of course it will need a lot of additional works, the other finding aids will be soon presented, however important question is still hanging: "are we ready for EAD?", especially for retrospective conversion? From the point of view of technology I think: Yes, we are. Of course the tools like XMetaL or XMLSpy, programs for automation of conversion from database to EAD template and from template to HTML would be helpful. As for archival side, two important questions have to be answered: do we have legacy inventories shaped and structured in the way, which let us put them into EAD template and do we have staff able to work on it? Here one could adopt a solution from of British National Archives (former Public Records Office) solution – a small group working on EAD for all archives in A2A project<sup>11</sup>. As to the first question, I am really not sure – with all the respect to what they had done – did the former generations of archivists left us legacy, which without hard work could be encoded according to EAD standards. Old descriptive practice did not included many basic elements that are essential to a proper finding aid and were not so standardized as they are now thanks to ISAD(G). During many decades archival way of thinking was like mentioned above motto: "my archives are unique", unifying these unique "universes" seems like very real challenge and job: "Can we fix it?"

---

*Zamoyskiego MK 133 z Archiwum Głównego Akt Dawnych w Warszawie z lat 1587 - 1595*, opr. A. Kot, W. Krawczuk, M. Kulecki, A. Sokół, G. Spyrka, *Sumariusz Metryki Koronnej*. Seria Nowa, t. II, Kraków 2001.

<sup>9</sup> Web site of Library of Congress - <http://www.loc.gov/ead/ead.html>

<sup>10</sup> Web site of Central Archives of Historical Records – <http://www.archiwa.gov.pl/agad/pomoce/MKInw.xml>

<sup>11</sup> *A2A. Access to Archives. Report, April 2000 – March 2002*, A2A Central Team, PRO 2002.