

데이터 요소의 표준화

지금까지 국명, 지명에 관해서만 언급했지만 일반적으로 [데이터베이스]에서는 [데이터 요소]의 표준화가 긴요하다. 특히 고유명사의 표준화는 정말로 시급한 일이다. 인명(사람이름)의 표기도 아직 정해진 표준이 없으며, 국명·지명도 문제가 있으며, 기관명의 표기 및 약어표기도 빨리 정해야 할 문제들이다. 한국전력공사를 한국전력으로 할 것인지 한전으로 할 것인지...등 수많은 단체명에 관한 표기법, 약어표기법, 일련번호부여들의 표준화 작업을 해야 할 것이다.

40 여년전 고교생시절, 한창 6.25전쟁이 어수선할 때, 대구의 한 길거리에서 흘러나온 팝송이 하나 있었다. 그 노래의 일부를 아직도 기억한다.

Istambul was Constantinople,
Now it's Istambul not Constantinople
.....
Even old New York was once New
Amsterdam
Why they say like it, I can't say
People just like it better that way.

이것마저 제대로 기억하고 있는지는 모르겠다. 이 말은 지금의 [이스탄불]은 옛날의 [콘스탄티노플]이었고 [뉴욕]은 옛날의 [뉴암스텔담]이었다는 사실이 이 노래에서 알았다.

필자는 5-6년전에 [한글코드의 표준화]를 위한 회의참석차 [요르단]의 수도인 [암만]에 갔다온 일이 있다. 여기서 알아낸 사실인데 로마제국이 요르단을

정복하기 전까지는 수도이름이 [암만]이었는데 로마가 정복하고 나서 [암만]이란 이름을 없애고 총독의 이름인 [필라델피아]로 바꾸어 버렸다고 한다. 요르단 독립후 원래의 이름을 되찾았지만 미국 동부에 있는 도시이름이 [필라델피아]인걸 보면 남이 버린 이름을 되쓰고 있는게 아닌가 싶다.

서울만 해도 경성, 한양, 한성 등 몇번씩이나 이름이 바뀌었다는 것을 아는 사람은 다 안다. 서울의 세종로, 을지로, 을곡로, 퇴계로 등의 이름의 역사는 불과 50년인 것 같다.

국제표준화기구(ISO)에서 투표문서 한건이 당도하였다. 번호는 DIS 3166-3. 우리나라에서는 이 문서에 대해서 찬반투표를 해야 한다. 이 문건의 내용을 훑어보니 달라진 나라이름에 관한 표준화 문건이다. 전 국명과 현 국명의 대조표가 들어있다. 모두 25개국의 나라이름이 바뀌었다.

귀에 익은 국가명에 대해 몇가지 소개하면 1.버마 -> 미얀마 2.체코슬로바키아 -> 체코와 슬로바키아로 양분 3.다호메이 -> 베닌 4.독일민주공화국 -> 독일 5.소련 -> 13개 공화국으로 분리 6.어퍼볼타 -

부르키나파소 등으로 변경된 것이다. 그밖에도 19개국이 변경되었는데 필자는 들도 보지도 못한 국가들이다. 여기서의 찬성, 반대는 속결지연의 의미밖에 없는 셈이다. 따라서 찬성하는 수밖에 없을 것 같다.

ISO에서는 국명 뿐만 아니라 일반지명에 대해서도 속속 표준화를 진행하고 있다. 현재는 시도명까지 정해진 상태이지만 조만간에 시, 면, 동의 이름까지도 확정될 것으로 보인다. 이 정도까지 진행하기 위하여 국제표준화회의가 15회가 열렸는데 이때까지는 한국이 참여조차 하지 않았다. 금년에 16회회의가 개최되었는데 여기에 처음으로 한국대표가 참석해서 의견개진을 했다.

만약 이러한 회의를 무시하고 참석하지 않은 채 몇년만 계속된다면 [조어도]가 [센카쿠렛토]로 표기되어도, [독도]가 [타케지마]로 표기된 채 투표문서가 당도하지 않았을까? 깊이 반성해야 할 일이라고 생각한다. OECD회원국 국가로서 부끄럽기 이를데 없다.

지금까지 국명, 지명에 관해서만 언급했지만 일반적으로 [데이터베이스]에서는 [데이터 요소]의 표준화가 중요하다. 특히 고유명사의 표준화는 정말로 시급한 일이다. 인명(사람이름)의 표기도 아직 정해진 표준이 없으며, 국명, 지명도 문제가 있으며, 기관명의 표기 및 약어표기도 빨리 정해야 할 문제들이다. 한국전력공사를 한국전력으로 할 것인지 한전으로 할 것인지...등 수많은 단체명에 관한 표기법, 약어표기법, 일련번호부여들의 표준화 작업을 해야 할 것이다.

단행본, 정기간행물, 기술보고서, 음반악보 등의 고유번호 붙이기 규칙도 정립되어야 한다. 이들이 모두 [데이터베이스]속의 [데이터요소]로서 담겨져 있어야 유효한 [데이터베이스]가 될 것이다. 이러한 표준화가 없이 있는 그대로 입력하는 [데이터베이스]는 쓰레기 처리장(Garbage in, Garbage out)에 불과하다.

우리나라에서 [데이터요소]의 표준화를 위한 시도

가 과연 얼마나 있었는지를 묻고 싶다. 이렇게 지지부진한 원인중 하나가 [정부부처 이기주의]가 아닌가 싶다. 표준화해야 할 100가지 항목가운데 단 1가지가 해당부처의 의견과

다르다고 해서 전체를 반대하는 분위기에서는 국가 표준의 제정을 기대할 수가 없지 않을까 싶다.

당연히 정부부처별로 [표준화추진기구]가 있어야 하겠지만 과연 부처마다 그러한 추진기구가 있는가? 정보산업 분야만 해도 그렇다. 필자가 알기로도 표준화 추진단체가 5-6개처나 된다. 각각이 다른 표준화체제로 표준을 생산하고 있다. 그런데 KS가 따로 있고, KTS가 따로 있고, KIS가 따로 있고... 이들 모두가 국가표준으로 자처하고 있지만 과연 그런가?

표준화 추진을 위한 정부기구로서 공업진흥청이 있었는데 어찌된 셈인지, 중소기업청 산하의 국립품질기술원으로 축소하여 버린 것은 표준화를 후퇴시킨 결과를 가져왔다. 비록 그대로 있었다고 해도 표준화를 위한 예산 확보를 몇년째 실패하고 있는지 모르겠다. 예산확보에 자신이 없으면 가능한 부처로 업무를 이관하는 것이 마땅하지만 그러지 못한 것은 오히려 표준화를 지연시키고 있는 결과가 된다.

우리말로 된 데이터베이스를 산업화하기 위해서는 조속히 [데이터요소]의 표준화를 서둘러야 할 것이다. 아직도 쓰레기처리식 데이터베이스를 그냥 계속할 것인가를 묻고 싶다. 표준화 추진사업이 국가 장래를 위하여 얼마나 중요한 것인가를 아직도 이해 못하는 예산당국은 지금도 [고속철도]에 돈 쓸 생각을 하는 모양이다. [고속철도]예산의 10만분의 1만 표준화 추진에 사용하였더라도 오늘날과 같이 비참하지 않을 것이다. **DC**

