

해양생물다양성 메타DB 고품질 구축 연구

양성용* · 박대우*

*호서대학교 벤처전문대학원 IT응용기술학과

A Study on Construction of High Quality Marine-Biodiversity Metadata DB

Sung-Young Yang* · Dea-Woo Park*

*Dept. of, IT application technology, Hoseo Graduate School Of Venture

E-mail : *sykytg@naver.com · *prof1@paran.com

요 약

해양생물다양성 자원에 대한 국가 주권을 인정하는 해양생물다양성협약의 발효로 생물다양성 자원을 확보하기 위한 세계 각국의 경쟁이 치열한 상황이다. 현재 우리나라 각 기관 및 대학들이 보관하고 있는 해양생물자원 정보가 산재되어 있고, DB에 대한 관리 미흡으로 인하여 학술적, 산업적 활용이 어려운 상황임을 인지하여, 통합적으로 관리할 수 있는 정보체계가 필요한 시점이다. 본 논문에서는 해양생물다양성 자원 중 고품질 해양생물자원에 대한 현황을 분석한다. 그리고 해양생물다양성 데이터 고품질 확보 방안을 위한 DB 품질 오류율 산정 기준을 적용하여 메타DB 구축 방안을 제시한다. 본 연구는 향후 해양생물자원에 대한 국가전략수립에 기여할 것으로 기대한다.

키워드

생물다양성, 해양생물자원 정보, 메타DB, DB 구축

I. 서 론

1992년 6월, 생물다양성협약이 채택된 이후, 생태적, 경제적 중요성에 대한 인식이 높아지고 생물다양성의 확보를 국가자산으로 인정하고 평가하게 되었다. 또한 유전자원 접근 및 이익공유(ABS, Access to genetic resources and Benefit-Sharing)를 위한 ABS 국제의정서 채택('10. 10, 일본 나고야)[1]됨에 따라 세계 각국은 생물자원 선점을 위하여 국가차원의 경쟁에 돌입하게 되었으며, 생물다양성 자원이 바이오 경제시대의 미래자원으로 인식됨에 따라 자원에 대한 확보가 치열해지고 있는 추세이다.

특히 우리나라에 산재되어 있는 해양생물다양성 자원을 확보·관리하여 자국의 자원보호를 위한 국가대응체계를 강화해야 하는 시점이다. 따라서 해양생물다양성 자원을 정보화하고, 표준화하여 산업계·학계·연구계에서 실시간 스마트 정보를 활용할 수 있는 메타 DB에 대한 연구가 필요하다.

본 논문 연구에서는 생물다양성 자원 중 해양생물자원에 대한 현황 분류와 DB 품질 오류율 산정 기준을 적용한 고품질 메타DB 구축에 대한 분

석을 제시하고자 한다.

II. 관련연구

2.1 생물자원 메타DB 구축 현황

생물자원의 중요성을 인식하여 수집 및 개발에 많은 투자를 하고 있으나 전략적인 국가통합관리 및 활용이 부족한 상황이다. '08년 생명공학육성 시행계획에 의하면 교과부, 환경부, 국토해양부, 농식품부 등 관련 경제부처가 생물자원 연구에 687.8억 원을 투자하여 국내 토종생물을 매년 1,000여종에 100만점 이상 발굴 확보하고 있다. 하지만 발굴된 자원의 정보화가 미흡하여 자원의 활용성이 매우 떨어지고 있으며 국립중앙과학관이 약 117만건의 생물자원 표본정보, 산림청 국립수목원이 95만의 식물, 곤충표본정보를 가지고 있으며, 국토해양부 해양생물 9,000여종의 정보를 서비스하고 있지만 절대적인 정보의 양이 부족한 실정이다[2].

그림 1은 GBIF의 2010년도 “선진국 대비 생물자원 보유 현황”에 관한 그래프이다.

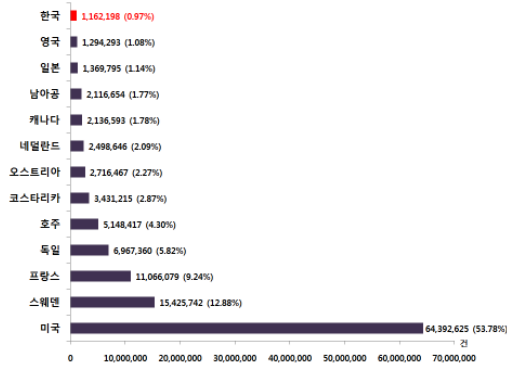


그림 1. 선진국 대비 생물자원정보 보유 현황

2.2 국가생물다양성정보통합시스템

국립중앙과학관 및 각 지방 자연사박물관 등 생물다양성의 정보화와 통합연계 시스템, 다양한 데이터 형태의 DB 구조를 관리하기 위하여 시스템을 구축하였으며, 국제기구와의 협력을 통한 생물다양성 정보의 국가경쟁력 확보함은 물론, 대국민이 참여할 수 있다. 생물다양성 포털사이트는 2004년부터 생물다양성정보를 구축하여 현재 약 116만건을 서비스하고 있다.

2.3 해양생물다양성 DB 및 시스템 현황



그림 2. 한국해양생물다양성정보시스템 화면

그림 2처럼 2007년 국내에서는 처음으로 우리나라에 서식하는 다양한 해양생물 정보를 제공하기 위한 해양생물의 종목별로 체계적으로 정리한 한국 해양생물다양성 정보시스템을 구축하여, KoMBIS에는 식물플랑크톤 2578종, 동물플랑크톤 201종, 염생식물 46종, 해조류 988종, 해양저서동물 5008종, 어류 977종, 해양포유류 31종 등 우리나라 18개 해양생물군에 속하는 총 9800여종의 생물명과 1800여개의 동종이명이 수록되어 있다.

특히 일부 종에 대해서는 상세정보, 출현정보, 분자마커정보 등도 수록되어 있으며, 정보를 수집한 참고문헌이 생물명과 연결되어 있어 사용자가 추가 정보를 확인할 수 있도록 했다[1].

KoMBIS는 또 학명이나 한글명을 통해 해양생물 종을 검색할 수 있으며, 그림 3처럼 분류체계에 따른 계통검색 기능을 이용하면 특정 분류군의 종목록을 볼 수 있도록 하는 등 종명을 쉽게 찾을 수 있도록 다양한 검색기능을 갖추고 있다. 정확한 종명이나 계통을 모르는 경우에도 일반어 검색기능으로 해당 단어가 포함된 종 목록을 찾을 수 있도록 했다.

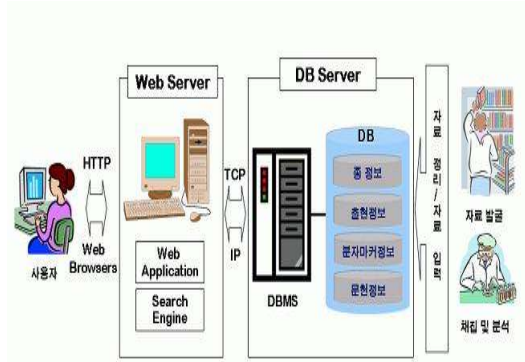


그림 3. 해양생물다양성 정보시스템 구성도

III. 해양생물자원 메타DB 구축 표준화 방안

각 부처별 업무 소관별로 구축되어 있는 해양생물자원의 정보화 자료의 통합 연계 시스템을 통한 표준화 전략을 수립하여 기관에 적합한 데이터 생산을 유도하고 활발한 데이터 교류가 필요하다.

해양생물자원 메타DB 구축 표준화 방안으로 생물자원관련 국제기구 및 국외 선진 기관과의 협력을 통해 정보화 자료의 공유를 위한 표준화 체계의 수립해야 할 것이다.

또한 생물자원 국제기구와 기관 등에서 권장하는 GBIF의 Darwin core 2 방식을 준수하여 국제적으로 국가생물자원 정보화 자료를 공유할 수 있는 체계 구축하여 표준화 방안을 마련해야 한다.

3.1 해양생물자원 text 메타DB 구축 표준화 작업 공정

해양생물자원은 메타DB text 구축과 메타DB 이미지를 구축하며, 각 기관의 산재해 있는 메타DB 구축을 연계하기 위해서는 표준화 공정이 필요하다.

표 1의 메타 DB text 구축에 대한 업무내역 표준화처럼 메타DB text구축 절차는 대상자료 선정 →자료구축준비→text 작성 및 입력→검수단계를 걸쳐 작성하게 된다.

표1. 메타 DB text 구축에 대한 업무내역 표준화

단위 작업	상세공정	업무내역
준비 단계	대상자료 선정	구축 관련 자료 확보/정리 및 검토
	자료구축 준비	해양생물자원 DB 분류군 및 구축점수 확정 및 수집
		실물 DB자료에 대한 리스트 작성
		필드 양식 작성 구축대상 실물 자료의 리스트 및 라벨링 작업
자료입 력단계	텍스트 작성 및 입력	선정된 생물자원 실물의 기본정보 및 상세정보를 작업자가 1차 작성 기본정보 및 상세정보(생물종 정보 외 기타 정보)는 전문가의 검수를 거친 필드규정에 맞춰 입력
	1단계 검수 2단계 검수 3단계 검수	1단계 검수
2단계 검수		중단위(생물종 정보) 검수 오타자, 띄어쓰기, 맞춤법 검수
3단계 검수		전문위원이 입력된 실물 DB정보와 실제 소장 실물자료의 정보와 일치 여부 확인(웹서비스 전수 검수) 기본설명에 정확도, 상세설명에 충실도 검수

3.2 해양생물자원 메타DB image구축 표준화 작업
공정

해양생물자원 메타DB image구축 절차는 표 2
처럼 메타DB image 구축에 대한 업무내역 표준
화방안으로 대상자료선정→자료구축준비→image
촬영→image 보정→검수단계를 걸쳐 작성하게
된다.

표2. 메타 DB image 구축에 대한 업무내역 표
준화

단위 작업	상세공정	업무내역
준비 단계	대상자료 선정	구축 관련 자료 확보/정리 및 검토
	자료구축 준비	해양생물자원 DB 분류군 수집
실물 DB자료에 대한 리스트 작성		
구축대상 실물 리스트 및 라벨링 작업		
촬영 단계	이미지 촬영	실물 자료가 훼손되지 않도록 취급 촬영 시 누락 발생을 방지하도록 목록 확인 촬영 원본의 원형이 최대한 유지될 수 있도록 촬영
		보정 단계
검수 단계	1단계 검수 2단계 검수 3단계 검수 4단계 검수	여백 및 노이즈 제거 형식, 누락 이미지 점검
		촬영리스트와 이미지 라벨 일치 및 수량 검수 촬영 상태, 이미지 보정 상태 검수
		1단계 검수 결과에 대한 작업자 검수
		전문가에 의한 리스트와 이미지 일치성 검수 촬영 상태 품질 검수
4단계 검수	이미지 자료와 텍스트 자료 일치성 검수	

IV. 데이터 고품질 확보 방안을 위한 메타
DB 품질 오류를 산정 기준

본 논문에서는 해양생물자원의 데이터 고품질
확보를 위한 방안으로 메타 DB에 대한 오류율을
산정하여 제시함으로써 국내외 생물자원에 대한
가치를 인정받고, 산학연에서 활용할 수 있는 고
품질 메타 DB를 창출하고자 한다.

데이터 고품질 확보 방안을 위한 메타 DB 품질
오류율 산정 기준은 표 3처럼 메타 DB text 정보
오류율 산정기준을 통하여 메타 DB 품질을 확보
하여야 한다.

표 3. 메타 DB text 정보 오류율 산정기준

항목	내용
오류검사	필드속성 : 필드속성 오류, 주소정보 오류, 색인연결 오류, 기타
	중정보 : 오타자, 띄어쓰기, 기호입력, 색인연결 오류, 기타
오류율 산정	필드속성 오류율 = (전체 오류 필드 개수/전체 필드 개수)*100
	중정보 오류율 = ((전체 오타자 자소수+띄어쓰기오류 외 자수)/전체 자소수)*100

또한 데이터 고품질 image 확보 방안을 위한 메타 DB image 품질 오류율 산정 기준은 표 4처럼 메타 DB image 정보 오류율 산정기준을 통하여 메타 DB 품질을 확보하여야 한다.

표 4. 이미지정보 오류율 산정기준

항목	내용
오류검사	원본비교, 해상도/크기, 노이즈, 여백, 파일형식
오류율산정	{(전체 원본비교+전체 파일형식 오류수)*0.2 + (해상도,크기+노이즈+여백+기타 오류수)*0.15}/전체 건수 *100

표 5에 나타난 결과와 같이 메타DB 구축 표준화 방안을 통해 고품질의 메타 text DB와 메타 DB image에 본 논문에서 제시한 기준을 적용시켜서 99%의 고품질 해양생물자원 메타DB 구축을 확보하게 되었다.

표 5. 메타DB 구축 표준화 방안 前後 오류율

구분	해양생물자원	
	메타DB text	메타DB image
샘플링수	1,000건	1,000건
대상생물자원	어류	어류
표준화 前 오류율	92.3%	87%
표준화 後 오류율	99.1%	99.3%

V. 결 론

우리나라에 분포하고 있는 생명자원은 약 10만 종에 이르며, 현재 밝혀져 있는 종은 3만종에 불과하다. 즉 7만종은 아직까지 밝혀지지 않고 있으며 이를 확보하기 위해 국가, 연구소 등에서 많은 노력을 하고 있다. 그러나 특히 해양생물 자원쪽은 다른 생물자원에 비해 표준 및 자원확보가 미비한 상태이다. 본 논문은 이를 위해 언제 어디서든지 활용할 수 있는 클라우드 컴퓨팅을 이용하여 표준화가 이루어져야 한다고 생각하였고 text, 이미지 데이터 표준화를 위한 구축방안과 데이터 구조의 필드요소를 분석하였다.

향후 연구로는 우리나라 실정에 맞는 해양생물 자원의 데이터 필드 구조에 대한 연구와 해양생물자원을 활용할 수 있는 클라우드 컴퓨팅 서비스에 대한 다양한 연구가 필요하다.

참고문헌

- [1] 박수진, 생물다양성협약 ABS 국제레짐의 논의 동향과 향후 일정, 해양정책연구부
- [2] 전병선, 조류지식네트워크를 위한 한국조류정보시스템 2010.08.
- [3] 임정희, 오히목, 박용하, 박종화, 감창배, 백운기, 생물다양성정보의 오늘과 내일, 세계생물다양성기구 한국사무국, 2009.11.
- [4] 백운기, 안승락, 박형선, 임정희, 이상명, 김동희, 민홍기, 생물종 기준표본의 디지털이미지 매뉴얼, 세계생물다양성정보기구, 2010.03.
- [5] 안부영외 2인, 생물다양성 학습을 위한 생물다양성 DB 활용에 대한 연구
- [6] 임정희, 생물다양성 정보체제 구축과 활용에 대한 연구, 2009.