GBIF Global Biodiversity Information Facility

Plinian Core Mentoring: strengthening best practices for mobilizing species information

FINAL ACTIVITY REPORT

Contents

1. Executive summary

This project sought to enhance capacity among the nodes currently working on creating best practices toward the mobilization of species information through the use of the Plinian Core Standard. Keeping in mind that the mentoring and capacity building is an ongoing process among the nodes, this year the working group met in Mexico City in February 2015 to work on technical issues regarding the standard and continuing building a mentoring network to support and achieve our goals as a working group. In this final report we include the activities we have completed during this year, including the meeting in Mexico, along the main technical and capacity achievements, and the current situation of the project after this year of continuous mentoring. It is important to note that all the objectives for the mentoring were completed, and that most of the final recommendations include further work needed to make sure the standard is working to fulfill our expectations.

2. Contact information

Patricia Koleff
Coordinator
CONABIO, Liga Periférico - Insurgentes Sur, Núm. 4903, Col. Parques del Pedregal, Delegación Tlalpan, 14010, México, D.F.
pkoleff@conabio.gob.mx
+ 52 55 5004 5005

3. Project summary

The Plinian Core Standard aims to facilitate the mobilization of species level information by standardizing and structuring information from different data-bases

through the use of common standards, including DwC and EML. The integration with these standards will allow the interoperability of species level information to the IPT and GBIF data portal with different species cataloguers that implement Plinian Core. In order to complete the implementation of the Plinian Core standard, the working group is holding a mentoring including all the nodes. As part of this mentoring, we continuously meet via Skype conferences, participate in online discussions, interact in small technical interchanges, and in February 2015 met in Mexico City to enhance individual capacities and contribute, at the same time, to the consolidation of a common standard, associate methodologies, and tools.

3.1. Activities completed

- Fine-tuned the Plinian Core Standard Abstract model (v3.2), and improved its related documentation.
- Documented best practices needed to describe and publish species level information in the context of GBIF and TDWG.
- Revised a number of controlled vocabularies used by the Standard.
- Updated and consolidated the work towards having a Plinian Core extension for GBIF's IPT.
    - See: http://tools.gbif.org/dwca-validator/extensions.do [under "Under Development"]
- Tested and refined the current concept mapping of Plinian Core-DwC and Plinian Core- EOL Schema.
- Planned the activities of the TDWG interest group in Species information and a task group around species level information standards.
    - See: http://www.tdwg.org/activities/species-information/charter/
- Documented the InvasivenessAtomizedClass

3.2. Ongoing and post-project activities

- Continue rrevising a number of controlled vocabularies used by the Standard.
- Finish the XSLT style sheets to transform Plinian Core xml files into DwC and EOL xml files.
- Continue planning activities of the TDWG interest group in Species information and a task group around species level information standards

4. Project objectives

The objectives of the mentoring program were:

- To fine-tune Plinian Core Standard Abstract model and improve its related documentation
- To document the best practices needed to correctly describe and publish species level information and make these available in the global context (GBIF, TDWG, etc.)
- To revise and finalize a number of controlled vocabularies used in the Standard
- To gain critical mass in publishing species information datasets in GBIF and EOL
- To update and consolidate a Plinian Core extension for GBIF's IPT
- To test and refine current concept mappings of Plinian Core-DwC and Plinian Core- EOL Schema.

- To contribute to the activities of the TDWG interest group in Species information and set in motion a task group around species level information standards.

5. Project deliverables

1. A new version of Plinian Core Abstract Model (it would be v3.2)
   - https://github.com/PlinianCore
   - https://github.com/PlinianCore/Documentation/wiki
   - https://github.com/PlinianCore/Documentation/wiki/PlinianCore_Terms
2. A "Best practices" document (attached)
3. A set of controlled vocabularies for some key terms
   - https://github.com/PlinianCore/Documentation/wiki/PlinianCore_Terms
4. Pilot datasets containing species level information published via GBIF and EOL in which the new elements of the standard and associate procedures are tested and fine-tuned (attached)
5. A new IPT extension for Plinian Core
   - The new IPT extension for Plinian Core
     http://www.gbif.es:8080/TallerIPT/resource.do?r=prueba_pliniancore321_final
     http://tools.gbif.org/dwca-validator/extensions.do
   - Abstract models
     https://github.com/PlinianCore/Sources/tree/master/xsd/abstract%20models/stable%20version
   - Completed APs
     SIB Colombia
     https://raw.githubusercontent.com/PlinianCore/Sources/master/xsd/application%20profiles/stable%20version//PliC3.2.1_AP_SIB-COLOMBIA.xsd
     CONABIO
     https://raw.githubusercontent.com/PlinianCore/Sources/master/xsd/application%20profiles/stable%20version/PliC3.2.1_AP_CONABIO.xsd
     MAGRAMA
     https://raw.githubusercontent.com/PlinianCore/Sources/master/xsd/application%20profiles/stable%20version/PliC3.2.1_AP_MAGRAMA.xsd
     INBIO
     https://raw.githubusercontent.com/PlinianCore/Sources/master/xsd/application+profiles/stable+version/PliC3.2.1_AP_INBIO.xsd
     GBIF.ES
     https://raw.githubusercontent.com/PlinianCore/Sources/master/xsd/application+profiles/stable+version/PliC3.2.1_AP_GBIF_ES.xsd
6. XSLT style sheets to transform Plinian Core xml files into DwC and EOL xml files
   As until very recent we were still working in the final version of the data base, and it was not sensible to work on an XSLT without a final version as that would have implied extra effort from the development team, we still have not been able to finish the development of the XSLT. Moreover, the IT people involved in that effort are no longer working for their nodes, so we are currently in the midst of reorganizing the development to finish this deliverable.
7. An active TDWG interest group in Species

- See: http://www.tdwg.org/activities/species-information/charter/

6. Evaluation: findings and conclusions

This mentoring exercise has strengthened our abilities and our vision on how to better work towards mobilizing species information, as well as to how to use best practices (in terms of documenting and using controlled vocabularies) to document our own information. During this year and thanks to the continuous interaction among the nodes fostered by the mentoring, we made significant advances in finishing and revising the abstract model of the Standard, thee APs for five different GBIF nodes. We continue to work together as a group, with constant feedback from all the nodes, and frequent communication via Skype.

7. Recommendations and lessons learned

After all the hard work of this year mentoring, we were able to learn how different nodes approach the topic on how to standardize and publish species level information and applied this knowledge to each individual case. Among all participants, we were able to discuss how to standardize different requirements, how careful practice and documentation (including the use of common controlled vocabulary for common needs) can lead to better practices, and how much we can advance in technical details when we work together as a group. We still have a lot of work to do, including making tests migrating real data, and finishing the XSLT style sheets to transform Plinian Core xml files into DwC and EOL xml files. This mentoring has showed us how much work can be accomplished by getting together for a period of time, and how important it is to maintain periodic meetings with the nodes in order to accomplish our goals.

We include here the answers to the survey in which all participants expressed their view of the most significant experiences gained during the onsite mentoring.

Development and IT, GBIF Spain

- Released a new version of Plinian Core 3.2 with several substantial and important changes such as the modification in the design that allows the user to use the Atomized and Unstructured elements in the same sequence.
- Analyzed the compatibilities between Plinian Core and EOL standards.
- Analyzed the procedures to create a XSLT process that will allow us to migrate Plinian Core to different standards.
- Approved the migration of the repository of Plinian Core on Google Code to GitHub.
- Analyzed the interoperability between different application profiles, so that in the future we are able to share the species pages among different portals.
- Learnt how other groups approach the topic of how to standardize and publish species information and apply this knowledge to our case

Angélica Cervantes, Liliana Lara, Esther Quintero, and Patricia Koleff, CONABIO and GBIF Mexico

- Documented best practices document for sharing species information
- Discussed the importance of the use of controlled vocabularies (atomized data) and the control of information in extenso (ancillary data).
- Mapped the controlled vocabularies in the Plinian Core Standard Abstract model.
- Discussed adapting Conabio's and other countries 'rich content species pages to the Standard
- Analyzed the interoperability among different application profiles
- Analyzed and enriched the compatibility of PLIC with other standards (i.e. EOL)
- For Liliana Lara and Angélica Cervantes: getting to know the Plinian Core Standard, the architecture of the system, its interoperability, and the capacity of the system to migrate information to other standards.
- The ongoing mentoring has strengthen our abilities and our vision on how to better work towards mobilizing species information, as well as to how to use best practices to document our own information.

José Trinidad Mendoza, Juan Kasa: CONABIO´s informatics department and GBIF node manager

- Adapted the AP to generate records of species following the new Plinian Core Standard Abstract model (v3.2).
- Worked in an XSLT process that will allow us to migrate and adapt the developments created by CONABIO to other standards.
- Contributed to the improvement of the documentation for the standard.
- Generated a compatible xml file following PLIC 3.2 in an automated way.
- The mentoring has allowed us to fine tune our skills and to exchange experiences with other IT managers from different nodes, which have broadened our community.

Carmen Quesada Ochoa: Herbarium University of Granada, Spain

- Developed a new version of the standard (Plinian Core 3.2) much more stable and suitable.
- Improved definitions and documentation of the concepts of the standard.
- Reviewed the use of controlled vocabularies based on the experience of CONABIO, in order to establish recommended vocabularies.
- Approved the use of Github documentation (https://github.com/PlinianCore)
- Analyzed and confronted the different APs (comparing concept to concept) to assess their compatibility.
- Analyzed and approved the use of EML GMP (dataset branch) metadata.
- Decided to focus in the TDWG new website to make sure we appear as a Working Group.
- In my case, as Herbarium of the University of Granada, the analysis, comparison and improvement of application profiles already developed, has

been essential as a starting point and for the acquisition of good practices in order to the development of a new specific application profile for our necessities (Algae).

Danny Vélez: Sistema de Informação sobre a Biodiversidade Brasileira

- Analyzed the possibility to create a PliC application profile for SiBBr and to use it in projects of species-fact-sheets documentation in Brazil.
- In Brazil the SiBBr is working with the Ministry of Science and Technology (MCTI), the Institute Chico Mendes de Conservação da Biodiversidade (ICMBIO), and the World Conservation Monitoring Centre (WCMC) to develop a system to create and visualize species-fact-sheet and It is intended that this system use the Plinian Core standard. The onsite February meeting was an excellent opportunity to see and share experiences in how other countries and teams are working in the implementation of Plinian Core.

Maria Mora: National Biodiversity Institute of Costa Rica (INBio) and GBIF CR.

- Improved PliC concepts definition using the information requirements of institutions that are generating or integrating species information. This ensures that the concepts included are needed to avoid overfilling the proposal.
- Assessed, homologated, and extended institutional application profile definitions that could serve as case examples for new PLiC users.
- Strengthened the PLIC community of developers by exchanging experiences among its members.
- Improved PLiC documentation.
- Shared design ideas to develop software tools to manage species records.
- Assessed controlled vocabularies to be included as part of PLiC recommendations.
- For INBio the mentoring has meant a great contribution to the implementation of the new data portal of the Costa Rican Biodiversity Information System (CRBio) based on the Atlas of Living Australia (ALA). The experience has allowed INBio to evaluate how other institutions in the Iberoamerican Community are integrating species records to enrich its experience. INBio in collaboration with GBIF.es, is currently developing a module for integrating species information based on an application profile specified after the workshop in Mexico.

Valentina Grajales: Software developer - SIB Colombia

- Improvements of the PlinianCore abstract model and update the version to 3.2.
- Advanced in the construction of the final XML schema for PlinianCore Abstract Model (latest version 3.2) and the validation process for the application profile that we are going to use in Colombia.
- Worked in a draft for SIB Colombia´s application profile.

- Agreed in the controlled vocabulary for a final schema that will be used to validate our data models.
- Validated a schema for the abstract model version 3.2, which we can now use to validate our application profile and xml data files.
- Worked on a group of assets to do transformations among data models using concept mappings between standards (for example DarwinCore extensions and EOL schemas).
- Worked on the tools to translate and transform PlinianCore to GBIF IPT extension and EOL schema.
- Shared new experiences, ideas, and best practices about how to improve the process to get, process, administer, and publish data about species.
- Proposed to use OneSky as a cloud solution to help in the translation process.
- As a developer for SIB Colombia, the analysis and share of ideas will help us with the development of our own practices and tools for new software applications.

Camila Plata: SIB Colombia

The Plinian core Mentoring held in February was crucial to the development of current activities of the Biodiversity information System of Colombia (SiB Colombia) in order to release the next version of the Biodiversity Catalog of Colombia . Plinian Core Abstract model (3.2) sequential structure now supports our principal need of having both atomized and unstructured information in order to reach a wide public, from children to scientist and decision makers. Additionally these version´s elements have a stronger biological background and their documentation is better supported; several elements are now repetitive which allows expressing the complexity of biological process such as migration and invasion. After several discussions in previous meetings, the group decided to include molecular data into the standard, taking advantage of strong initiatives such as GenBank and the international barcode of life (IBOL).

Starting from the abstract model we were able to establish SiB Colombia's final application profile upon which we are now building the information architecture of the new version of the Catalog. During the mentoring the application profiles of participant institutions were tested in order to find incompatibilities among them and to document best practices to ensure interoperability between the different profiles, this test improved our own profile, and ensured that we will be able to share species information successfully.

In order to improve the collective work around the standard we are now employing informatics tools such as OneSky for translations, and GitHub to host the code. We hope these tools will make collaborative work easier.

8. Future plans

- Make tests migrating real data to the Standard (Conabio Is about to start inputing data into its data base integrated according to the standard)
- Propose a draft for an AP for Algae

- Work in the appropriate software architecture to properly share data of species between entities; guarantee the interoperability between different application models.
- Study mechanisms to insert species-fact-sheets using the IPT PliC extension in the processes to produce Data Papers.
- Preliminary statistics of the number of projects, institutions, and GBIF nodes and participants that implement Plinian Core (long term).
- Incorporate the Plinian Core Standard in CONABIO´s new species website "enciclovida", which will allow thorough and more specific searches to atomized information at the species level (http://bios.conabio.gob.mx/)

In case you want to learn more, please visit:

https://en.wikipedia.org/wiki/Plinian_Core

or the TDWG species interest group at:

http://www.tdwg.org/activities/species-information/charter/

For further comments please contact:
Dr. Patricia Koleff
CONABIO, Liga Periférico - Insurgentes Sur, Núm. 4903, Col. Parques del Pedregal, Delegación Tlalpan, 14010, México, D.F.
pkoleff@conabio.gob.mx
+ 52 55 5004 5005

9. Signature of the project main contact point

Februrary 2, 2016

Dra. Patricia Koleff Osorio

Conabio, Head of the Mexican Delegation and project coordinator