Data-driven model (DDM) for collection development and management: from library data to institutional value generation

Cristian Alejandro Chisaba Pereira
Collection Development Librarian, Biblioteca General Alfonso Borrero Cabal, S.J., Pontificia Universidad Javeriana (Bogotá, Colombia).

Abstract: DDM for collection development and management is a proposal for capture, analyze, discover insights from *data* and make proposals from Collection Development and Management to the library of the Pontificia Universidad Javeriana. Our model finds a value for the operation and innovation of the library from the data present in the information systems, repositories, social networks, ERP, CRM, among others. Exploiting the data and designing value proposals for new services or products from the library. Finally, the DDM identifies the profiles for an *analytical team* for the collection development and management.
Keywords: data-driven model, collection development, collection management, library data, analytics

1. Introduction

The data is one of the most relevant active of an organization or institution. Millions and millions of data are present in different software like CRM, ERP, KMS, BI, SGA and of course, ILS, LPS, repositories, proxy software, among others. A lot of data source appear every week in our organization or institutions: social network, web sites, smartphones, apps, software of IoT. The big technologic companies like Google, Amazon and Netflix have built their success around data (Vicelich, 2018). The importance around data is a paradigm that will change the world and our libraries.

This situation builds up a possibility for make new ways to manage our libraries. Also will help to create, design and test operative services or products associate to data.

Data-driven model for collection development and management take advantage of this context for build an opportunity in our libraries. DDM propose a framework for make collection development and management based on data of

Received: 14.1.2021 Accepted: 25.6.2021 ISSN 2241-1925

our bibliographic resources. We'll introduce skills and competences for an Analytical Team, which is the proposed for DDM. Finally, we are going to explain how DDM will transform the collection development and management in the library Alfonso Borrero Cabal, S.J. of the Pontificia Universidad Javeriana.

2. Data ecosystem at the library

When William S. Cleveland (2001) spoke about new role of statistics centered on data analyst and the term "data science" became popular. It would begin a great theoretical, conscious and practical development of data in modern society (Roy, 2018). Data is changing the world and our libraries are starting to make new activities around data. We think that the term data is becoming strength because data science has integrated different discipline like statistics, mathematics, information science, computer science among others.

We can find several examples of data science in libraries as an initiative about data management, data visualization, data curation, data analysis, data quality, etc. In North Carolina State University, New York University Libraries and Arizona State University "have been supporting researchers with training in computational and statistical skills" (Oliver, 2019). Jiang Wu, Jingxuan Cai, Miao Jin, & Ke Dong (2018) analyzed how libraries can expand its services, including data services, for help scientific funding application or, in another scenario, text mining has becoming a "modern knowledge management" in libraries (Qing Zhu, Yiqiong Wu, Yuze Li, Jing Han, & Xiaoyang Zhou, 2018). Also, we find projects that take advantage of data for analize Hakka (China) genealogical migration (Cheng, 2018), create a model of Big data smart library (Simović, 2018) or even Scientific Intelligence Service at Pontificia Universidad Javeriana Library (PUJ, 2019).

We can realize that data is building new capabilities and ~~its~~ open a panorama of opportunities for our libraries. Analyze data, explore it, capture it, preserve it, diagnose and discover insights for predict future behaviors and build new strengths around our libraries.

3. Data-driven Model for Collection Development and Management

When we start to think and design Data-driven Model for Collection Development and Management was important to analyze if our work was not delivering value to our Institution. So, knowing what those problems were, we analyzed how we were making our work. One service that we offer to faculties is Analyses of Bibliographic Resources from syllabus. We realized that for analyze about 1.863 bibliographic resources between books, journals, articles, and others from Faculty of Psychology. We were spending about 12 months and found that at least 16% were not in the library and therefore should be purchased. It was problematic for our Institution because the majority of

students were active and all bibliographic resource are crucial for the academic and pedagogic process.

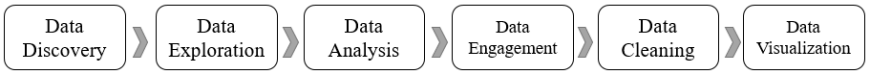
With these problems and other like reprocessing, late delivery of reports and wasted information and data available in the library, we start at 2019 to design the DDM. For this, we follow the next steps:

1. Identify problems or opportunities around Collection Development and Management.
2. Analyze how it worked the Information Analysis and Bibliographic Purchase sections. Because they were working in different process around Collection Development and Management like selection, acquisition, preservation, purchase, cataloguing, etc.
3. Design the first version of DDM.
4. Implement the first version of DDM.
5. Design a new version of DDM.
6. Communicate DDM for library and interested party into the Pontificia Universidad Javeriana.

We *Identify problems or opportunities* like reprocessing, late delivery of reports and wasted information and data available in the library because we did know that we were not deliver value to University and its was related with. For example, Information Analysis Section were working only with cataloguing, description and preservation of bibliographic resources and Bibliographic Purchase Sections were working only purchasing resources bibliographic resources. We realized that process like evaluation, assessment or conservation were being excluded from their operation. Then, with this problematic, we think a new way to make the things better and we found in data an opportunity for *Design the first version of DDM*.

In general, DDM always considered 4 clear dimensions:

We realized that data in libraries have diversity and different sources like Information Library System, Open Journal System, Dspace, social networks, data from site web, data from EzProxy. Then, Collection Development and Management process are fundamentals, they are the core of this model: we selection, evaluate, assess, but with data like the active key for make a new Collection Development and Management. Data exploitation is all the skills, competencies, process or methodologies for use data and convert it in insights for our library and of course, for our Institution. But we define the next process key for our model:



1. *Data discovery*: It means that we identify when, where or how is data.
2. *Data exploration*: We search, into the data that we have found. It´s important to identify patrons, qualities, quantities and a general view of what kind of data we have. Use software like R, Python or even Excel help us to understand the data.
3. *Data analysis*: we look if data make sense and answers our problem. Deepen with new question or reformulating previous ones.
4. *Data engagement*: Identify if external sources of data (outside library or even our Institution) can complement our data or answers the problem or questions.
5. *Data cleaning*: we build concepts and arguments with data. It´s important to verify if, really, our data responds to our problems. Use software like R, Python or other.
6. *Data visualization*: finally, we design reports, views, infographics or a visualization like a storytelling data approach.

In this way, we design our first version of DDM:

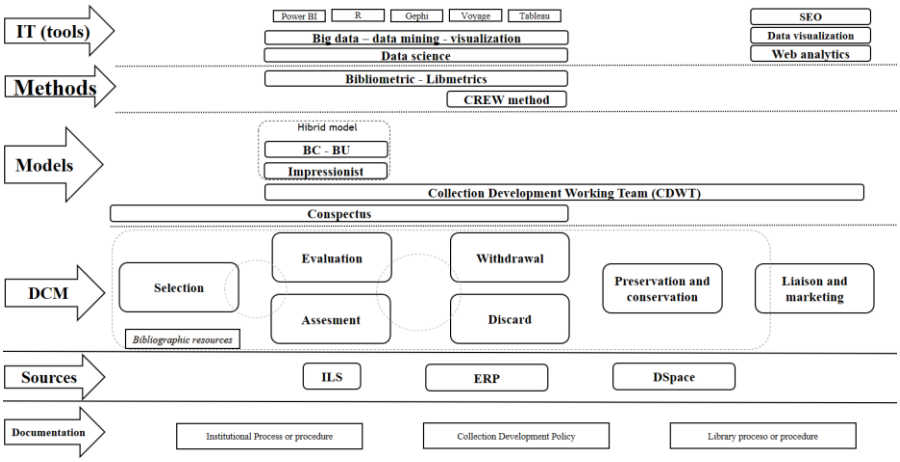


Figure 1. DDM version one.

Afterwards, we *Implement the first version of DDM* in the same, Analyses of Bibliographic Resources from syllabus, but, this time with Faculty of Engineering. Previously, with Faculty of Psychology we had spent 12 months analyzing 1.863 bibliographic resources. When use DDM approach with Faculty of Engineering we found that we had spent 1 month analyzing 4.337 bibliographic resources. That means that before DDM we spent 220 days real of

work analyzing data of 1.863 bibliographic resources, with DDM we could analyzed the same quantity in just 17 days.

Once we implement the first version of DDM, we iterate with another bibliographic resources data and we build a new version with an *operative model approach.*

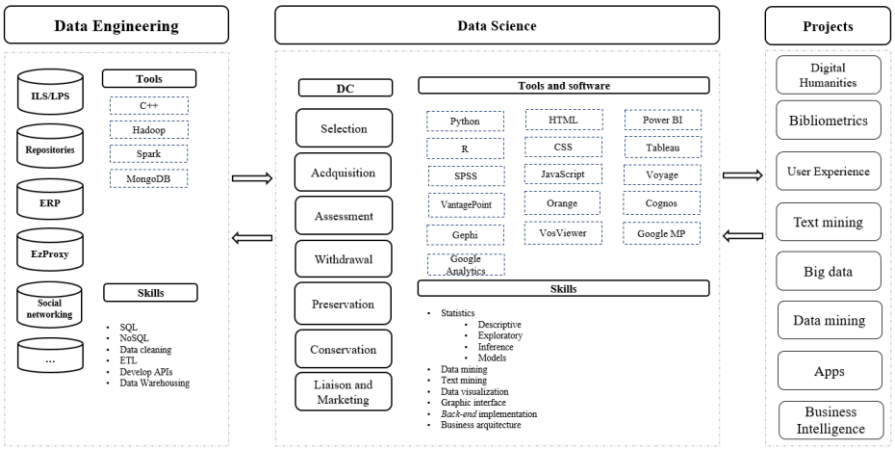
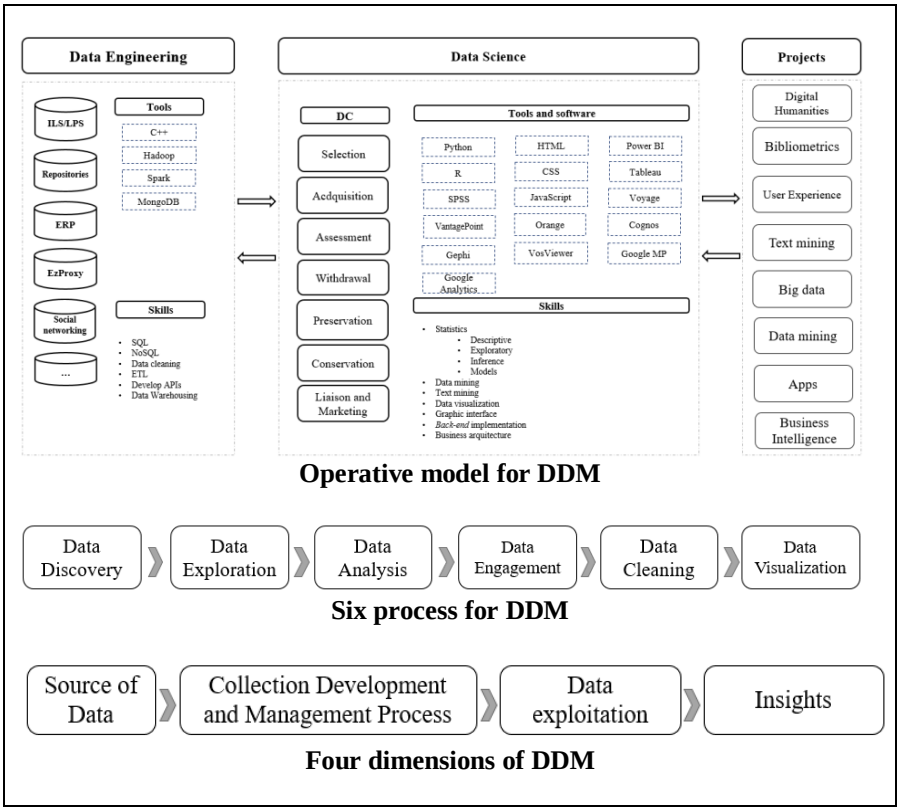


Figure 2. DDM version two.

In Figure 2, DDM is a mature model and we propose to operate with a Data Engineering, a focus from Data Science (here, Analytics Team is acting) and finally oriented to projects like Digital Humanities, Big Data, Data and Text Mining, among others.

The last version has motivated new initiatives into library. For example, for 2020, we proposed that Collection Development Section modified its processes and create new skills and capabilities for the people who work in our section. It will be described in the next item.



Data-driven Model for Collection Development and Management

4. Skills and competencies for an Analytics Team for DDM

The people who make DDM is the key for a success model. Different skills are required like technical, business knowledge, management project and soft skills, among other things. An approximation about the skills and competencies for an Analytical Team for DDM is closely linked to data skills. Semeler, Pinto & Rozados (2019) suggest that librarians "should more about data management and analysis by acquiring some data scientist skills such as those of a hacker, scientist, quantitative analyst, trusted adviser, or an expert business person.". Also, McCaffrey & Giesbrecht (2016) sketch out that some skills and competencies for a data librarian are very important like *data management and curation, data and geospatial visualization* and advanced data reference services.

Furthermore, Federer (2018) define the competencies and skills for a data librarianship., For *data management*, is important data management planning, preservation, curation or stewardship; support for data use and data analysis and, of course, development data services. Skills for *Information technology* like data

visualization, programming (R, Python, SPSS, MATLAB, Gephi, Geographic Information software, etc.) and finally another skills and competencies like teaching and instruction, marketing and outreach, library skills and soft skills. The Alianza Caoba, a Colombian alliance between universities and best companies in Colombia, for develop projects of data science and data analytics wrote a document called Perfil Alianza Caoba – Reporte Técnico that describe four dimensions for an analytic team: *Business and domain, Information Technology, Analytics* and *management* (Alianza Caoba, 2017).

This context is relevant because an Analytics Team for DDM must integrate a holistic skills and competencies for projects and challenges into library and around education ecosystem. We conclude the next dimensions for an Analytics Team for DDM:

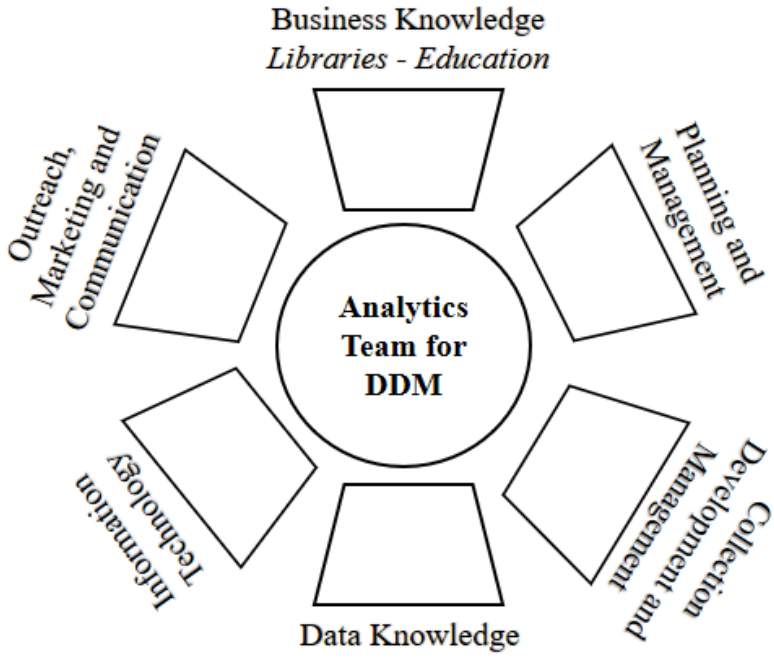


Figure 3. Dimensions for Analytics Team for DDM

DDM Analytics Team is an important and relevant element for success. This six dimensions are: Business Knowledge, Planning and Management, Collection Development and Management, Data Skills, Information Technology Skills and Outreach, Marketing and Communication:

- *Business Knowledge*: the key skills or competencies for this dimensions are:

  - Trends in education: methodologies, business model, training, etc.
  - Business model.
  - Profile of institutions or universities.
  - Virtual or blended learning.
  - Development services and products for business.
  - Business Intelligence.
- *Planning and Management:* the key skills or competencies for this dimensions are:
  - Planning projects.
  - Planning projects in Business Analytics.
  - Manage risk, teams, resources, technologies, time, communications.
  - Different methodologies: agile, design of services or products.
  - Connect with institutional projects
- *Collection Development and Management*: the key skills or competencies for this dimensions are:
  - Selection, acquisition, evaluation and assessment, withdrawal, preservation and outreach.
  - Planning projects in Collection Development and Management.
  - Design of services or products.
  - Trends in Collection Development.
  - Technology trends
- *Data Knowledge*: the key skills or competencies for this dimensions are:
  - Data management.
  - Data curation.
  - Data visualization.
  - Data quality.
  - Design of data services or data services.
  - Analyze information and data.
  - Connect with institutional data services.
  - Business Analytics.
- *Information Technology*: the key skills or competencies for this dimensions are:
  - Extract, Transform and Load data.
  - Data stewardship
  - Programming: R, Python, Java, etc.
  - Databases management.
  - APIs development.
  - Data warehousing
- *Outreach, Marketing and Communication*: the key skills or competencies for this dimensions are:
  - Soft skills: communication, negotiation, empathy, leadership, innovation, opening to the people, creativity.

- Marketing: digital, social media, SEO, email, outbound, etc.
- Information visualization.
- Connect with business: CXO, institutions, strategic people, operations people, tactic people, creative people, etc.
- Be creative! Be passionate!

In this way, DDM integrate an Analytics Team with diverse skills and competencies that let to our libraries improve the decision making, create value for his institution or ecosystem, create new services or products, making better users in library for a better society.

5. DDM in Biblioteca Alfonso Borrero Cabal, S.J., Pontificia Universidad Javeriana:

During 2019, DDM allowed to the Alfonso Borrero Cabal, S.J. Library change way it worked. Our library has four sections called Specialized Services, Fundamentals Services, Information Analysis and Bibliographic Purchase. When we create DDM, Information Analysis and Bibliographic Purchase sections operated together all collection development and management process like selection, acquisition, preservation, purchase, cataloguing, etc. DDM enabled the process re-design of Information Analysis Section and Bibliographic Purchase Sections as a result of a new sections called Development Collection and eliminated Bibliographic Purchase Sections. This new sections were formally created in May 2019 with a corpus of function fundamentals in collection development and management, but with a new approach based on data.

Ending 2019, the Alfonso Borrero Cabal, S.J. Library had satisfactory results as a result of the DDM. Development Collection Section analyzed all bibliographic data of source of information available in syllables of Pontificia Universidad Javeriana. We start analyzing bibliographic data of 10 (ten) departments with an amount of 7.129 source of information. We analyzing data about loans, inventories, thematic areas, who uses the source of information, frequency of use, comparison with another source of information and one of the most important results: how much does it cost to acquire new sources of information or how much would it cost to predict a budget for the future purchase of bibliographic resources. This point is important for our University. Until 2019, we estimate a same budget with an increase between 5% to 10%, year by year. With this DDM approach, we realize, and different departments, that, it amounts of budget is insufficient for the bibliographic resources of the syllables, after DDM approach our University will need increase budget in 300% in the next year or find new strategies for offer resources like books, journals, data bases, videos, among others.

Finally, for 2020, Alfonso Borrero Cabal, S.J. Library will consolidate DDM in Collection Development Section with interest of strengthening the Analytics Team and offering predictive and prospective reports for the institution, which

allows strategic decisions and design future services or products from our Library.

6. Conclusions

For 18 months, we went on to carry out a development of traditional collections and we started to develop collections based on the strengthening of data and new methodologies of the technological context. Also, DDM approach has allowed us:

- Optimize our operation in Development Collection Section.
- Improve response time to our University even more than 50% compared to operations without the DDM approach.
- Articulate us with projects that were previously isolated in our Library. One project called Data Mining Project has allowed to align library needs in relation with functions and responsibilities in Data Engineering according to Operative Model for DDM.
- Historically, when Bibliographic Purchase Sections was operating just offered to the University purchase of bibliographic resources. Since 2019, with DDM approach our faculties and department have begun to increase their applications in the face of the management of their bibliographic collections, future budgets and the creation of strategies for the use of all library services.
- Qualify our librarian team with new skills and capabilities around data, its management, analysis, visualization, etc.
- Accompany the process of create bibliography for new academic programs and define strategies, budgets for this bibliography.

Librarians are exploring new capabilities around library data. For Pontificia Universidad Javeriana Library, data has become in a fundamental active, and of course, has changed how we are working and how we will give to our Institution new services or products. For us, DDM like an approach from Collection Development Section has been a possibility to redesign our processes to deliver better results and anticipate to the academic, administrative and research needs of Pontificia Universidad Javeriana. From Collection Development Section, we have been building a new paradigm around data. Data has allowed us renovate our skills because the Analytical Team needs to know and to do for discover insights from data.

References

Alianza Caoba. (2017). *Perfil Alianza Caoba – reporte técnico.* Universidad de los Andes: Bogotá. Retrieval from https://minciencias.gov.co/sites/default/files/upload/convocatoria/anexo_no._1_-_perfil_citizen_data_scientist_caoba.pdf

Chang, C. C. (2018). Hakka genealogical migration analysis enhancement using big data on library services. *Library Hi Tech, 36*(3), 426–442.

Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review, 69*(1), 21–26. doi:10.1111/insr.2001.69.issue-1

Federer L. (2018). Defining data librarianship: a survey of competencies, skills, and training. *Journal of the Medical Library Association : JMLA*, 106(3), 294–303. doi:10.5195/jmla.2018.306

Jiang Wu, Jingxuan Cai, Miao Jin, & Ke Dong. (2018). Embedding funding consultation in library services : A co-occurrences network analysis of knowledge flow in scientific funding. *Library Hi Tech, 36*(3), 378–399. https://doi-org/10.1108/LHT-06-2017-0127

Knaflic, C. N. (2015). *Storytelling with data. a data visualization guide for business professionals.* Wiley.

Loriene Roy & Dan Sholler. (2019). What Reference Librarians Should Know about Data Science, *The Reference Librarian*, 60:2, 109-116, DOI: 10.1080/02763877.2018.1560964

McCaffrey, M, Giesbrecht, W (2016) Teaching data librarianship to LIS students. In: Kellam, L, Thompson, K (eds) *Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago: Association of College and Research Library, pp. 355–373.

Pontificia Universidad Javeriana. (2019). *Biblioteca Alfonso Borrero Cabal, S.J.* Bogotá. Retrieval from https://www.javeriana.edu.co/biblos

Pontificia Universidad Javeriana. (2019). *Servicio de Inteligencia Científica en la Biblioteca de la Pontificia Universidad Javeriana.* Bogotá. Retrieval from https://www.javeriana.edu.co/biblos/inteligencia-cientifica

Semeler, A. R., Pinto, A. L., & Rozados, H. B. F. (2019). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*, 51(3), 771–780. https://doi.org/10.1177/0961000617742465

Simović, A. (2018). A Big Data smart library recommender system for an educational institution. *Library Hi Tech, 36*(3), 498–523. https://doi-org/10.1108/LHT-06-2017-0131

Vicelich, C. (2018). How data became the secret sauce to commercial and customer success in the 21st century. *CIO (13284045)*, 6.