

# Bioinformatics for Researchers in Life Sciences: Tools and Learning Resources

Author: Natalia S. Rodriguez Muxica  
Editor: Rafael Anta

Institutions for  
Development Sector  
Competitiveness, Technology,  
and Innovation Division

DATABASE/DATASET  
IDB-DT-66

February 2022

# Bioinformatics for Researchers in Life Sciences: Tools and Learning Resources

Author: Natalia S. Rodriguez Muxica  
Editor: Rafael Anta

<http://www.iadb.org>

Copyright © [2022] Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Note that the link provided above includes additional terms and conditions of the license.

The results offered in this database/dataset are those compiled by the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



---

## **“Bioinformatics for Researchers in Life Sciences: Tools and Learning Resources”**

---

***Methodology Report***

***November 2021***

***Natalia S. Rodriguez Muxica  
Rafael Anta***

---

### **INTRODUCTION**

The experience of the COVID-19 pandemic has shown that bioinformatics, the intersection of biology and computer science, has a fundamental role in scientific research strategies in all the disciplines involved in fighting COVID-19: epidemiology, virology, biomedicine, genomics and synthetic biology, among others. This experience also suggests that bioinformatics skills in Latin America are incipient and that its scientific systems cannot take full advantage of the tools and data available thanks to the rise of open science and collaboration through global networks. In this context, CTI commissioned a study to produce a catalog of bioinformatics tools and related training that may help stimulate and encourage the development of bioinformatics skills in the region.

### **OBJECTIVES**

#### **General Objective**

Investigate which are the main digital skills required in bioinformatics to work in the response to the COVID-19 pandemic, and analyze the availability of resources for the development of these skills.

## **Specific Objectives**

1. Identify the main software and database tools used in the scientific disciplines of epidemiology, virology, biomedicine, and synthetic biology; and its associated costs.
2. Identify on the inventory of software tools and databases which digital skills are basic for any scientist and which digital skills are advanced.
3. Identify the availability of resources to learn how to use these tools and databases, and associated costs.
4. Identify skills necessary for the development of new bioinformatics applications, and the adoption of bioinformatics applications developed by others.

## **METHODOLOGY**

The project “*Bioinformatics skills to face the COVID-19 Pandemic: Tools and Learning Resources*” was executed between June 2020 and October 2021 in four main phases:

1. Search and identification of relevant software for researchers in life sciences
2. Search and identification of relevant learning resources in bioinformatics
3. Review and enrichment of both catalogs based on feedback
4. Identification of skills necessary for the development of new bioinformatics applications, and for the adoption of bioinformatics applications by researchers from Latin America.

### **Search and identification of relevant software for researchers in life sciences**

The tools gathered in this catalog are derived from three main sources: previously known tools by the author of the study, recommendations by researchers from Latin America, online inquiries both in spanish and in english, and based on the tools recommended or taught as part of the curriculum of the learning resources gathered in the second catalog.

Based on an initial exploratory search for tools and similar catalogs, the methodology and catalog template were established, defining the information needed from each tool, the categories used to classify them, and the selection criteria. All tools included in the final catalog were required to be available through the internet at the time this project was finished (either as a web-based tool or downloadable), to be accessible either with or without any charge, and finally, they were not required to be actively supported or updated as long as they continued to be functional.

The following are the variables used in the catalog of bioinformatic tools:

- Tool\_Name: Brand name or frequently used name for the tool.
- ID: Identification number for reference and internal use.
- Main\_URL: Uniform resource locator (web address) to the tool or platform or to its provider.
- Usage: Main function of the tool or intended usage of the platform.
  - Categories:
    - Data Analysis: tools dedicated to all kinds of data analysis or to support analytical functions.
    - Visualization: tools whose main function is to generate visual representations and graph data.
    - Repository and DB: tools and organizations dedicated to providing access to digital information and/or physical resources.
    - Data Storage: tools dedicated to facilitating the storage of research data.
    - Communication and Community: tools dedicated to foster communication within the scientific community.
    - Marketplace and Collaboration: tools and platforms that facilitate access to resources, services and for peer collaboration.
    - Lab and Resource Management: tools for internal management of inventories, resources, records and other laboratory management activities.
    - Other: for tools that do not fit into one of the previous categories, or that collect several tools at once.
- Biotech\_Fields: Fields within the biotechnology and life science sector that make use of the tool or that benefit directly from these resources.
  - Categories:
    - All
    - Biomathematics/Bioinformatics/Computational Biology
    - Biostatistics
    - Genetic Engineering
    - Genomics
    - Transcriptomics
    - Proteomics
    - Metabolomics
    - Cell Biology
    - Molecular Biology
    - Microbiology
    - Biochemistry
    - Biophysics
    - Physiology
    - Neurobiology
    - Immunology
    - Pharmacology
    - Epidemiology
    - Oncology
    - Biomedicine

- Zoology
  - Marine Biology
  - Ecology
  - Evolution
  - Environmental Biology
  - Chemistry
  - Systems Biology
- Description: Detailed description and complementary information about the tool or platform.  
\*This information has been collected from various sources and edited as needed.
- Technology\_Type: Type of technological platform used to provide the tool.
  - Categories:
    - Web based/Cloud App: for platforms accessible on-line.
    - Desktop/Mobile App: for platforms with a downloadable desktop or mobile application option.
    - Programming Environment/Library: for tools provided as downloadable scripts or open-access libraries.
    - Other: for non-identified tools.
- Level\_of\_Complexity: "Combined level of complexity of the tool or platform itself and of the required knowledge expected to effectively handle it. This is an arbitrary distinction made to distinguish tools that require general information and skills available for most life science undergrads (tagged as ""Basic""), from those tools that require either further programming and computational skills or specialized knowledge in a field (tagged as ""Advanced""). Tools with both ""Basic"" and ""Advanced"" levels of complexity have been tagged as ""Basic to Advanced"". The tools where the level of complexity was not clear were also included in the ""Basic to Advanced"" category."
  - Categories:
    - Basic
    - Advanced
    - Basic to Advanced
- Pricing\_Model: Main pricing model type of the tool or platform.
  - Categories:
    - Free / Open Source
    - Freemium
    - Subscription only
    - Usage based
    - Base plus overage
    - One-time payment
    - Other pricing model
- Pricing\_Detail: "Detailed information about the cost of accessing the tool or platform and possible conditions applied according to the user's sector (academia or industry) or other factors. \*Values as of June 2021. "

- **Documentation\_and\_Tutorials:** Supplemental information about the tool functionalities, installation manuals, usage guides and other tutorials, for those cases where it was available.
- **Alternative\_URLs:** Alternative sources for related scientific publications or secondary platforms that provide access to the tool.
- **Notes:** Supplemental notes and comments for potential users.
- **Year\_Published:** Year the tool was published. Based on their registered date in scientific publications, official website information or alternative sources such as source code repositories.
- **Country\_or\_Region:** Country or region where the tool was developed, based on authors affiliations and/or a company's country of operations.

### **Search and identification of relevant learning resources in bioinformatics**

The learning resources gathered in this catalog are derived from three main sources: previously known training opportunities by the author of the study, from tutorials and training recommended by the tool's official websites, general online inquiries both in spanish and in english, and direct consultation in online learning platforms or relevant communities: the International Society for Computational Biology (ISCB), iGEM, Coursera, EdX, Udemy, European Molecular Biology Laboratory (EMBL), Swiss Institute of Bioinformatics (SIB), and open courseware from universities, among others.

As well as for the catalog of tools, based on the initial exploratory search the methodology and catalog template were established, defining the information needed from each learning resource, the categories used to classify them, and the selection criteria. All resources included in the final catalog were required to be active or imparted between the period of 2020 and 2021, to be available either online, presentially or in a hybrid format, and to be accessible either with or without any charge.

As a criteria the learning and training opportunities considered were those that did not require significant technical background or a level of education higher than a bachelor degree, and that were shorter and simpler than specialized undergraduate, graduate, and postdoctoral studies, and, finally, those resources that were accessible for life science professionals without requiring full time dedication for long periods of time and that could be complimentary with their current job responsibilities.

The following are the variables used in the catalog of bioinformatics learning resources:

- **Resource\_Name:** Name of the program.
- **ID:** Identification number for reference and internal use.



- Institution\_and/or\_Platform: Institution or platform that provides the particular program.
- Main\_URL: Main URL to access the program itself or the information to register.
- Resource\_Type: Type of learning resource according to the amount of instructive hours, depth of the content delivered and format (written, video, interactive, etc.).
  - Categories:
    - Diploma
    - MOOC/Specialization
    - Course
    - Workshop
    - Webinar/Video Tutorial
    - Manual/Written tutorial
    - Other
- Modality: Modality used to deliver the contents.
  - Categories:
    - Online
    - Presential
    - Hybrid
- Level: Depth of the content being delivered through the program.
  - Categories:
    - Basic usage information
    - Introductory to the field/topic
    - Advanced training
- Content\_Details: Details about the content and curriculum of the program, as described by the institution or platform providing it. For resources in Spanish the original language has been maintained. (This information has been collected from various sources and edited as needed).
- Availability: The frequency format in which the program is delivered.
  - Categories:
    - Sporadic
    - Timed (yearly/monthly/etc)
    - On demand
- Date\_or\_Duration: Date when the program is being imparted, for those that are sporadic or timed with a certain frequency. Duration of the program, for those cases that are delivered on demand.
- Language: Languages available for the program, either spoken or with the support of subtitles.
  - Categories:
    - Español
    - Português
    - English

- Cost: Official cost of the program. \*Provided in the local currencies, and as of June 2021.
- Costs\_Details: Details about the cost breakdown, exceptions, preferential rates and other information when needed.
- Biotech\_Fields: Fields within the biotechnology and life science sector that make use of the tool or that benefit directly from these resources.
  - Categories:
    - All
    - Biomathematics/Bioinformatics/Computational Biology
    - Biostatistics
    - Genetic Engineering
    - Genomics
    - Transcriptomics
    - Proteomics
    - Metabolomics
    - Cell Biology
    - Molecular Biology
    - Microbiology
    - Biochemistry
    - Biophysics
    - Physiology
    - Neurobiology
    - Immunology
    - Pharmacology
    - Epidemiology
    - Oncology
    - Biomedicine
    - Zoology
    - Marine Biology
    - Ecology
    - Evolution
    - Environmental Biology
    - Chemistry
    - Systems Biology
- Tools\_Addressed: References to the tools in the "Repository of Digital Tools for Researchers in Life Sciences" that are used during the program or that are related to the contents imparted according to the curriculum.
- Addittional\_URLs: Alternative sources for related information about the course or other courses provided by the same organization.

### **Review and enrichment of both catalogs based on feedback**

This phase consisted in iterative reviews to incorporate, refine and update all the information required based on the formats defined previously for both catalogs.

## **Identification of skills necessary for the development of new bioinformatics applications, and for the adoption of bioinformatics applications by researchers from Latin America**

This final phase consisted first in a preliminary analysis of the tools and learning resources gathered in the previous steps in order to identify valuable insights, challenges and opportunities regarding the availability of digital tools as well as the availability and accessibility of training opportunities for them to be used by life science professionals in Latin America. This analysis also included the identification of tools used in the fields of greatest interest for the study: Epidemiology, Virology, Biomedicine, Synthetic Biology (Genetic Engineering & Systems Biology); and the identification of basic and advanced bioinformatic skills. The aforementioned analysis used as selection criteria the intersection of the two variables “Biotech Fields of Relevance” and “Level of Complexity”, considering as a basic skill the use of those tools classified both as “Biotech Fields of Relevance” equal to “All” and “Level of Complexity” equal to “Basic”. This criterion allows us to consider tools that can be useful for all fields of life sciences and that in turn do not require advanced digital skills (such as programming languages, use of bash/unix, etc.) or advanced knowledge of a particular discipline.

Secondly, a focus group was conducted in order to further explore the skills necessary to the development of new bioinformatic applications by professionals from Latin America, as well as to understand the current trends and challenges in this field inside the region. For this purpose, a group of bioinformatic professionals from the Latin American RSG-ISCB network was invited to participate in a discussion session on the 7th of June 2021, where a total of 11 members participated, representing 7 countries from the region (Chile, Peru, Colombia, Argentina, Mexico, Venezuela, and Ecuador). All insights gathered from this instance were summarized in the final presentation titled “*Final Results IDB Project: Bioinformatic Skills - 2021 - Convergencia BID \_EN.pdf*”.

## **RESULTS**

The project “*Bioinformatics skills to face the COVID-19 Pandemic: Tools and Learning Resources*” resulted in two main products:

5. A Catalog of Software for Researchers in Life Sciences
6. A Catalog of Learning Resources in Bioinformatics for Researchers in Life Sciences

Both results of the study can be found as part of the final package released by the Inter-American Development Bank in December 2021, together with this report, under the titles

*“Bioinformatics-Tools-for-LatAm-1-Catalog-of-Digital-Tools-for-Researchers-in-Life-Sciences-v4.0”* and *“Bioinformatics-Tools-for-LatAm-2-Catalog-of-Learning-Resources-in-Bioinformatics-v2.0.csv”*. A final presentation of all results and insights derived from the study can be found in the same release package under the title *“Final Results IDB Project\_ Bioinformatic Skills - 2021 - Convergencia BID\_EN.pdf”*.

## **ACKNOWLEDGEMENTS**

The current project was commissioned by the Inter-American Development Bank, under the department of Institutions for Development and its Competitiveness Technology and Innovation Division. All content was created between June 2020 – October 2021, and published in December 2021. Author: Natalia S. Rodriguez Muxica; editor: Rafael Anta.



## Metadata Sheet

---

Field	Description
<b>Dataset Title</b>	[EN] Bioinformatics Tools: A Catalog of Software for Researchers in Life Sciences. [ES] Herramientas Bioinformáticas: Un catálogo de software para investigadores en ciencias de la vida.
<b>Description</b>	<p>The experience of the COVID-19 pandemic has shown that bioinformatics, the intersection of biology and computer science, has a fundamental role in scientific research strategies in all the disciplines involved in fighting COVID-19: epidemiology, virology, biomedicine, genomics and synthetic biology, among others. This experience also suggests that bioinformatics skills in Latin America are incipient and that its scientific systems cannot take full advantage of the tools and data available thanks to the rise of open science and collaboration through global networks. In this context, CTI commissioned a study to produce a catalog of bioinformatics tools and related training that may help stimulate and encourage the development of bioinformatics skills in the region.</p> <p>The dataset is a catalog of bioinformatics software for researchers and professionals working in life sciences. It includes tools for varied uses such as: data analysis, visualization, repositories and databases, data storage services, scientific communication, marketplace and collaboration, and lab resource management. Most tools are available as Web-based or desktop applications, while others are programming libraries. This catalogue includes more than 300 different tools, together with 10+ suggested entries for other third-party repositories that could be of use.</p>
<b>Dataset Type</b>	Tables in CSV files
<b>Publisher/Editor</b>	Rafael Anta Competitiveness, Technology and Innovation Division (CTI) Institutions for Development (IFD)
<b>Creator Contact Email</b>	Natalia S. Rodriguez Muxica, research consultant, <a href="mailto:norodrig@uc.cl">norodrig@uc.cl</a> Rafael Anta, <a href="mailto:rafaela@iadb.org">rafaela@iadb.org</a>
<b>Creation Date</b>	June 2020 – October 2021
<b>Time Period Covered</b>	2020 – 2021
<b>Country (s)</b>	Global

<b>Background Document Attachments</b>	N/A
<b>Variables Names and brief description</b>	<ul style="list-style-type: none"> <li>- Tool_Name: Brand name or frequently used name for the tool.</li> <li>- ID: Identification number for reference and internal use.</li> <li>- Main_URL: Uniform resource locator (web address) to the tool or platform or to its provider.</li> <li>- Usage: Main function of the tool or intended usage of the platform. <ul style="list-style-type: none"> <li>o Categories: <ul style="list-style-type: none"> <li>• Data Analysis: tools dedicated to all kinds of data analysis or to support analytical functions.</li> <li>• Visualization: tools whose main function is to generate visual representations and graph data.</li> <li>• Repository and DB: tools and organizations dedicated to providing access to digital information and/or physical resources.</li> <li>• Data Storage: tools dedicated to facilitating the storage of research data.</li> <li>• Communication and Community: tools dedicated to foster communication within the scientific community.</li> <li>• Marketplace and Collaboration: tools and platforms that facilitate access to resources, services and for peer collaboration.</li> <li>• Lab and Resource Management: tools for internal management of inventories, resources, records and other laboratory management activities.</li> <li>• Other: for tools that do not fit into one of the previous categories, or that collect several tools at once.</li> </ul> </li> </ul> </li> <li>- Biotech_Fields: Fields within the biotechnology and life science sector that make use of the tool or that benefit directly from these resources. <ul style="list-style-type: none"> <li>o Categories: <ul style="list-style-type: none"> <li>• All</li> <li>• Biomathematics/ Bioinformatics/ Computational Biology</li> <li>• Biostatistics</li> <li>• Genetic Engineering</li> <li>• Genomics</li> <li>• Transcriptomics</li> <li>• Proteomics</li> <li>• Metabolomics</li> <li>• Cell Biology</li> <li>• Molecular Biology</li> <li>• Microbiology</li> <li>• Biochemistry</li> <li>• Biophysics</li> <li>• Physiology</li> <li>• Neurobiology</li> </ul> </li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>● Immunology</li> <li>● Pharmacology</li> <li>● Epidemiology</li> <li>● Oncology</li> <li>● Biomedicine</li> <li>● Zoology</li> <li>● Marine Biology</li> <li>● Ecology</li> <li>● Evolution</li> <li>● Environmental Biology</li> <li>● Chemistry</li> <li>● Systems Biology</li> </ul> <ul style="list-style-type: none"> <li>- Description: Detailed description and complementary information about the tool or platform. *This information has been collected from various sources and edited as needed.</li> <li>- Technology_Type: Type of technological platform used to provide the tool. <ul style="list-style-type: none"> <li>○ Categories: <ul style="list-style-type: none"> <li>● Web based/Cloud App: for platforms accessible on-line.</li> <li>● Desktop/Mobile App: for platforms with a downloadable desktop or mobile application option.</li> <li>● Programming Environment/Library: for tools provided as downloadable scripts or open-access libraries.</li> <li>● Other: for non-identified tools.</li> </ul> </li> </ul> </li> <li>- Level_of_Complexity: "Combined level of complexity of the tool or platform itself and of the required knowledge expected to effectively handle it. This is an arbitrary distinction made to distinguish tools that require general information and skills available for most life science undergrads (tagged as ""Basic""), from those tools that require either further programing and computational skills or specialized knowledge in a field (tagged as ""Advanced"). Tools with ""Basic"" and ""Advanced"" levels of complexity have been tagged as ""Basic to Advanced"". *The tools where the level of complexity was not clear were also included in the ""Basic to Advanced"" category." <ul style="list-style-type: none"> <li>○ Categories: <ul style="list-style-type: none"> <li>● Basic</li> <li>● Advanced</li> <li>● Basic to Advanced</li> </ul> </li> </ul> </li> <li>- Pricing_Model: Main pricing model type of the tool or platform. <ul style="list-style-type: none"> <li>○ Categories: <ul style="list-style-type: none"> <li>● Free / Open Source</li> <li>● Freemium</li> <li>● Subscription only</li> <li>● Usage based</li> <li>● Base plus overage</li> <li>● One-time payment</li> </ul> </li> </ul> </li> </ul>
--	--

	<ul style="list-style-type: none"> <li>• Other pricing model</li> </ul> <ul style="list-style-type: none"> <li>- Pricing_Detail: "Detailed information about the cost of accessing the tool or platform and possible conditions applied according to the user's sector (academia or industry) or other factors. *Values as of June 2021. "</li> <li>- Documentation_and_Tutorials: Supplemental information about the tool functionalities, installation manuals, usage guides and other tutorials, for those cases where it was available.</li> <li>- Alternative_URLs: Alternative sources for related scientific publications or secondary platforms that provide access to the tool.</li> <li>- Notes: Supplemental notes and comments for potential users.</li> <li>- Year_Published: Year the tool was published. Based on their registered date in scientific publications, official website information or alternative sources such as source code repositories.</li> <li>- Country_or_Region: Country or region where the tool was developed, based on authors affiliations and/or a company's country of operations.</li> </ul>
<b>Number of variables</b>	15
<b>Copyright and CC Licenses</b>	The results offered in this dataset are those compiled by the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.
<b>Additional Notes</b>	N/A



## Metadata Sheet

---

Field	Description
<b>Dataset Title</b>	[EN] Bioinformatics Tools: A Catalog of Learning Resources for Researchers in Life Sciences. [ES] Herramientas Bioinformáticas: Un catálogo de recursos de aprendizaje para investigadores en ciencias de la vida.
<b>Description</b>	<p>The experience of the COVID-19 pandemic has shown that bioinformatics, the intersection of biology and computer science, has a fundamental role in scientific research strategies in all the disciplines involved in fighting COVID-19: epidemiology, virology, biomedicine, genomics and synthetic biology, among others. This experience also suggests that bioinformatics skills in Latin America are incipient and that its scientific systems cannot take full advantage of the tools and data available thanks to the rise of open science and collaboration through global networks. In this context, CTI commissioned a study to produce a catalog of bioinformatics tools and related training that may help stimulate and encourage the development of bioinformatics skills in the region.</p> <p>The catalog of learning resources in Bioinformatics is a dataset of available training opportunities to acquire the knowledge needed to make use of bioinformatic and digital tools for professionals working in the diverse fields of life science. These resources include training in diverse formats, ranging from diplomas and specializations, to courses and tutorials. The resources are available in three main languages: English, Spanish and Portuguese, and both in on-line and in-person modalities. This catalogue includes 144 different trainings, together with 5 suggested entries for other third-party repositories with training opportunities that could be of use.</p>
<b>Dataset Type</b>	Tables in CSV files
<b>Publisher/Editor</b>	Rafael Anta Competitiveness, Technology and Innovation Division (CTI) Institutions for Development (IFD)
<b>Creator Contact Email</b>	Natalia S. Rodriguez Muxica, research consultant, <a href="mailto:norodrig@uc.cl">norodrig@uc.cl</a> Rafael Anta, <a href="mailto:rafaela@iadb.org">rafaela@iadb.org</a>
<b>Creation Date</b>	June 2020 – October 2021
<b>Time Period Covered</b>	2020 – 2021

<b>Country (s)</b>	Global
<b>Background Document Attachments</b>	N/A
<b>Variables Names and brief description</b>	<ul style="list-style-type: none"> <li>- Resource_Name: Name of the program.</li> <li>- ID: Identification number for reference and internal use.</li> <li>- Institution_and/or_Platform: Institution or platform that provides the particular program.</li> <li>- Main_URL: Main URL to access the program itself or the information to register.</li> <li>- Resource_Type: Type of learning resource according to the amount of instructive hours, depth of the content delivered and format (written, video, interactive, etc.). <ul style="list-style-type: none"> <li>o Categories: <ul style="list-style-type: none"> <li>• Diploma</li> <li>• MOOC/Specialization</li> <li>• Course</li> <li>• Workshop</li> <li>• Webinar/Video Tutorial</li> <li>• Manual/Written tutorial</li> <li>• Other</li> </ul> </li> </ul> </li> <li>- Modality: Modality used to deliver the contents. <ul style="list-style-type: none"> <li>o Categories: <ul style="list-style-type: none"> <li>• Online</li> <li>• Presential</li> <li>• Hybrid</li> </ul> </li> </ul> </li> <li>- Level: Depth of the content being delivered through the program. <ul style="list-style-type: none"> <li>o Categories: <ul style="list-style-type: none"> <li>• Basic usage information</li> <li>• Introductory to the field/topic</li> <li>• Advanced training</li> </ul> </li> </ul> </li> <li>- Content_Details: Details about the content and curriculum of the program, as described by the institution or platform providing it. For resources in Spanish the original language has been maintained. (This information has been collected from various sources and edited as needed).</li> <li>- Availability: The frequency format in which the program is delivered. <ul style="list-style-type: none"> <li>o Categories: <ul style="list-style-type: none"> <li>• Sporadic</li> <li>• Timed (yearly/monthly/etc)</li> <li>• On demand</li> </ul> </li> </ul> </li> <li>- Date_or_Duration: Date when the program is being imparted, for those that are sporadic or timed with a certain frequency. Duration of the program, for those cases that are delivered on demand.</li> </ul>

	<ul style="list-style-type: none"> <li>- Language: Languages available for the program, either spoken or with the support of subtitles. <ul style="list-style-type: none"> <li>o Categories: <ul style="list-style-type: none"> <li>• Español</li> <li>• Português</li> <li>• English</li> </ul> </li> </ul> </li> <li>- Cost: Official cost of the program. *Provided in the local currencies, and as of June 2021.</li> <li>- Costs_Details: Details about the cost breakdown, exceptions, preferential rates and other information when needed.</li> <li>- Biotech_Fields: Fields within the biotechnology and life science sector that make use of the tool or that benefit directly from these resources. <ul style="list-style-type: none"> <li>o Categories: <ul style="list-style-type: none"> <li>• All</li> <li>• Biomathematics/ Bioinformatics/ Computational Biology</li> <li>• Biostatistics</li> <li>• Genetic Engineering</li> <li>• Genomics</li> <li>• Transcriptomics</li> <li>• Proteomics</li> <li>• Metabolomics</li> <li>• Cell Biology</li> <li>• Molecular Biology</li> <li>• Microbiology</li> <li>• Biochemistry</li> <li>• Biophysics</li> <li>• Physiology</li> <li>• Neurobiology</li> <li>• Immunology</li> <li>• Pharmacology</li> <li>• Epidemiology</li> <li>• Oncology</li> <li>• Biomedicine</li> <li>• Zoology</li> <li>• Marine Biology</li> <li>• Ecology</li> <li>• Evolution</li> <li>• Environmental Biology</li> <li>• Chemistry</li> <li>• Systems Biology</li> </ul> </li> </ul> </li> <li>- Tools_Addressed: References to the tools in the "Repository of Digital Tools for Researchers in Life Sciences" that are used during the program or that are related to the contents imparted according to the curriculum.</li> <li>- Addittional_URLs: Alternative sources for related information about the course or other courses provided by the same organization.</li> </ul>
--	--

<b>Number of variables</b>	16
<b>Copyright and CC Licenses</b>	The results offered in this dataset are those compiled by the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.
<b>Additional Notes</b>	N/A



# Bioinformatics for Researchers in Life Sciences

## *Tools and Learning Resources*



Final Results

*IDB Convergence Platform*

***Natalia S. Rodriguez Muxica***

*IDB Consultant*

*June 8th 2021*

*\*Adaptation: October-November 2021*

# General Objective of the Study

Investigate which are the main digital skills required in bioinformatics to work in the response to the COVID-19 pandemic, and analyze the availability of resources for the development of these skills.

# First specific objective

- Identify the main software and database tools used in the scientific disciplines of epidemiology, virology, biomedicine, and synthetic biology; and its associated costs.

# Bioinformatics Tools Catalog

Bioinformatics Toolkit						
Bioinformatics Toolkit v1.0 Bioinformatics Toolkit v2.0 Bioinformatics Toolkit v3.0 Learning Resources						
VIEWSShareAutomationsApps						
By Usage (shared)7 hidden fieldsFilterGrouped by 1 fieldSorted by 2 fieldsColorShare view						
Find a view	Name	#	Type of Technology	Main Link	Alternative Links	Biotech Fields of Relevance
All	USAGE					
By Usage (shared) ✓	▶ Data Analysis	Count 160				
By Field*	USAGE					
Tool Suggestion Fo...	▶ Visualization	Count 25				
(For analysis: Basic...	USAGE					
(For analysis: Field...	▶ Repository & DB	Count 85				
(Internal use)	USAGE					
Create a view	▶ Data Storage	Count 6				
Grid +	USAGE					
Form +	▶ Communication & Con	Count 4				
Calendar +	USAGE					
Gallery +	▶ Marketplace & Collab	Count 3				
Kanban +	USAGE					
Gantt + Pro	▶ Lab & Resource Man	Count 15				
	USAGE					
	▶ Other	Count 20				



# Tools Catalog: *general statistics, by type of usage*

302

Bioinformatic Tools  
(+13 links to other repositories)

157

Data Analysis

85

Repository & DB

25

Visualization

15

Lab & Resource  
Management

6

Data Storage

4

Communication  
& Community

3

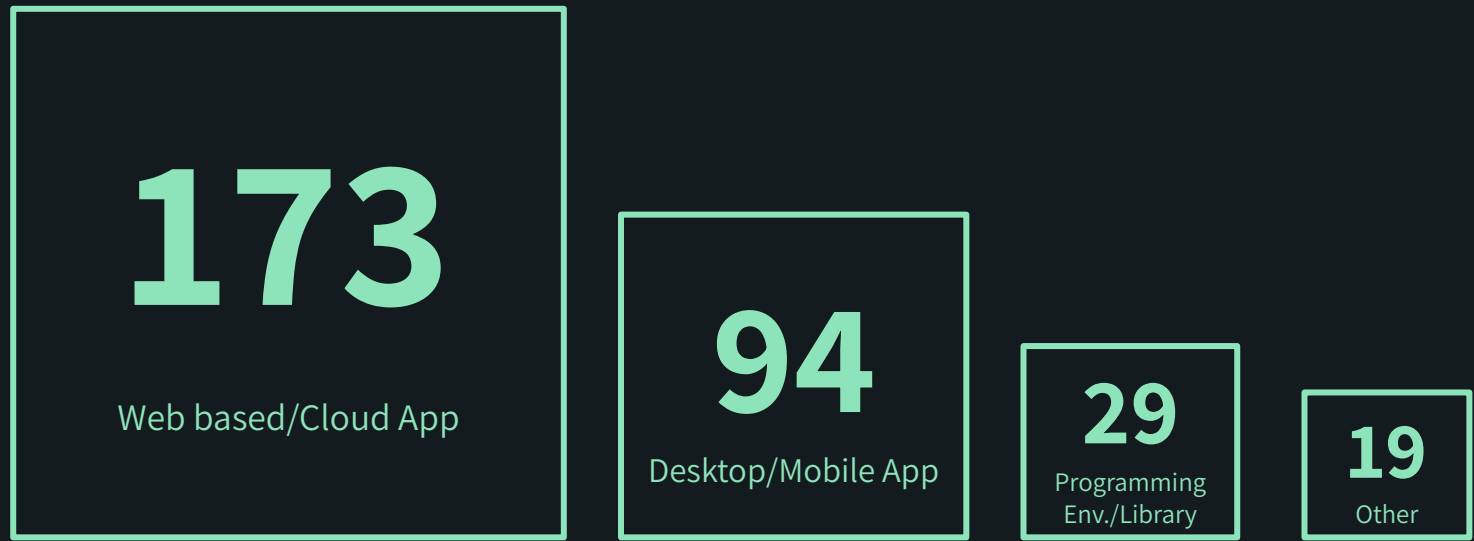
Marketplace &  
Collaboration

20

Other

Past-1989:	11 (4%)
1990-1999:	28 (9%)
2000-2009:	78 (25%)
2010-present:	124 (39%)
N/A:	74 (23%)

# Tools Catalog: *general statistics, by type of technology*



# Tools Catalog: *Fields of greatest interest for the study*

- **Tools by fields of interest:**

- Epidemiology → 18
- Virology = Microbiology → 65
- Biomedicine → 21
- Synthetic Biology = Genetic Engineering + Systems Biology → 187

# Tools Catalog: *Countries of origin* (\* raw data)

## Country of Origin

USA/UK

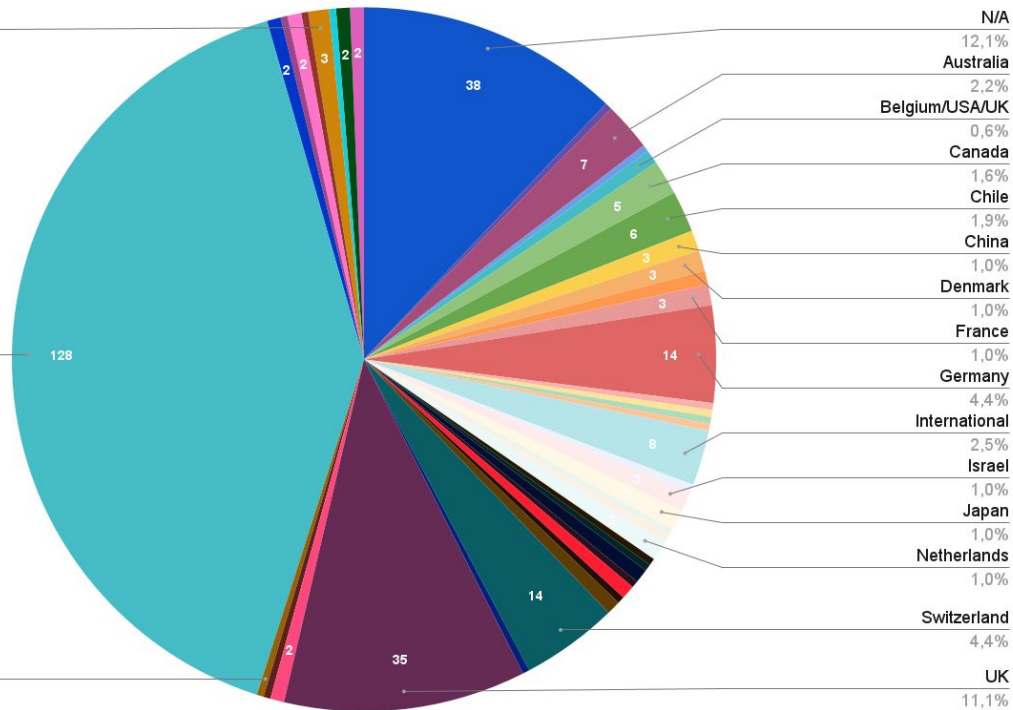
1,0%

USA

40,6%

UK,Switzerland,USA

0,3%



# Tools Catalog: *Zoom-in by countries*

Country Sub-groups	% (#)
<b>National Initiatives</b>	<b>69,5% (219)</b>
<b>US+</b>	<b>48,3% (152)</b>
<b>UK+</b>	<b>16,2%(51)</b>
<b>Switzerland+</b>	<b>6,0% (19)</b>
<b>Europe+</b>	<b>34,6% (109)</b>
<b>LATAM+</b>	<b>2,5% (8)</b>
<b>International Collaborations</b>	<b>10,8% (34)</b>



# Tools Catalog: *Associated costs*

## Pricing Model

**Other pricing model**

6,3%

**One time payment**

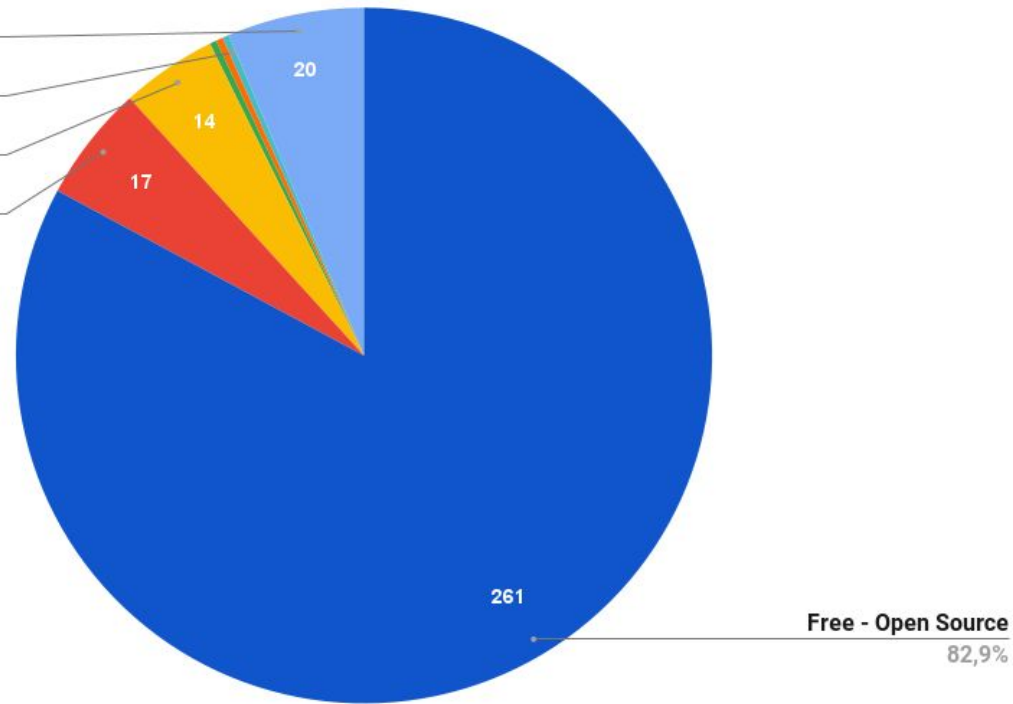
0,3%

**Subscription only**

4,4%

**Freemium**

5,4%



# Tools Catalog: *Associated costs*

**278**

**With free options: Free +  
Freemium**

**17**

**Payed tools**

**20**

**Variable or  
undetermined  
prices**

**Range in USD/year  
\$43-\$50400**

- Subscription only: 14
- Usage based: 1
- Base + overage: 1
- One time payment: 1

# Tools Catalog: *Associated costs, open-source challenges*

webact.org



This site can't be reached

The connection was reset.

Try:

- Checking the connection
- [Checking the proxy and the firewall](#)

ERR\_CONNECTION\_RESET

Details

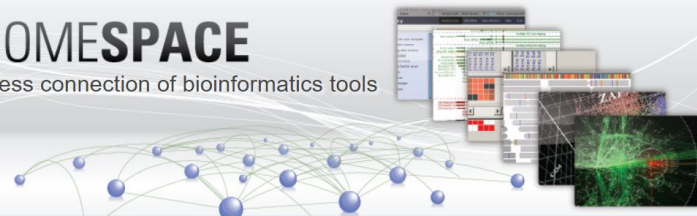
genomespace.org



[What is GenomeSpace?](#) [Tools](#) [Recipes](#) [Documentation](#) [Developers](#) [Support](#) [About](#)

## GENOMESPACE

Frictionless connection of bioinformatics tools



STATUS 11.18.19 06:02PM

**With the discontinuation of NHGRI funding for GenomeSpace we have shut down the servers.**

GenomeSpace Recipes can be found at <http://recipes.genomespace.org/> however data transfer through GenomeSpace will not be available.

More details can be found at <http://www.genomespace.org/news/>

[Calendar of Upcoming Events](#)



Tweets by @genomespace

 **GenomeSpace Team**  
@genomespace

The GenomeSpace project ends "tomorrow" November 15, 2019 due to expiration of its NHGRI funding. Please save any data from your

[Citing GenomeSpace](#)

To cite your use of GenomeSpace, please reference Qu K, Garamszegi S, Wu F, et al. [Nature Methods](#). 2016 Jan 18. doi: 10.1038/nmeth.3732.



## Second specific objective

- On the inventory of software tools and databases, identify which digital skills are basic for any scientist and which digital skills are advanced.

# Tools: *Basic v/s advanced skills*

- **Skills that could be considered basic:**

- → **73 tools** meet this criterion

- Criterion used:

- The tools considered as part of the “Basic Skills” were those in the set "Biotech Fields of Relevance"= “All” + “Level of Complexity” ="Basic"
- This criterion allows us to consider those tools that can be useful for all fields of life sciences and that in turn do not require advanced digital skills (such as programming languages, use of bash / unix, etc.) or advanced knowledge of a particular discipline.

- Some general basic skills to highlight:

- Big data analysis with programming tools in R or Python
- Use of user-friendly visualization and analysis tools: pymol / rasmol, Benchling / Snapgene
- Direct and systematic consultation of databases
- \* IP analysis for particular technologies or sectors
- \* Internal management: digital lab notebooks, image repositories, and complementary data management tools

# Tools: *Examples by field of interest*

- Tools by fields of interest:
  - Epidemiology → 18

fludb.org/brc/home.spg?decorator=influenza

**IRD Influenza Research Database** About Us Community Announcements Links Resources Support Workbench Sign In

SEARCH DATA ANALYZE & VISUALIZE WORKBENCH SUBMIT DATA HELP

### Search

Search our comprehensive database for:

- Sequences & Strains
- Animal Surveillance
- Sequence Feature Variant Types
- Immune epitope data
- 3D protein structures
- Host Factor Data
- Antiviral Drugs

Browse All Search Types

### Analyze

Analyze data online:

- Sequence Alignment
- Phylogenetic Tree
- Sequence Variation (SNP)
- Metadata-driven Comparative Analysis
- HPAI H5N1 Clade Classification
- Swine H1 Clade Classification
- BLAST

Browse All Tools

### Save to Workbench


Sign up for a workbench to:


- Store and share data
- Combine working sets
- Integrate your data with IRD data
- Store and share analyses
- Custom search alert


Sign Up! Sign In

### Supported Programs

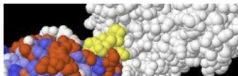
Click on a supported program of interest to go to program description page.

 **NIAID Functional Genomics**

 **NIAID Systems Biology**

 **NIAID CEIRS**

### Highlights



#### 3D Protein Structure

Visualize protein structures in 3D. Users can display sequence conservation score on a structure and highlight experimentally determined epitopes as well.

### What's New with Flu

Updated 13 May 2021

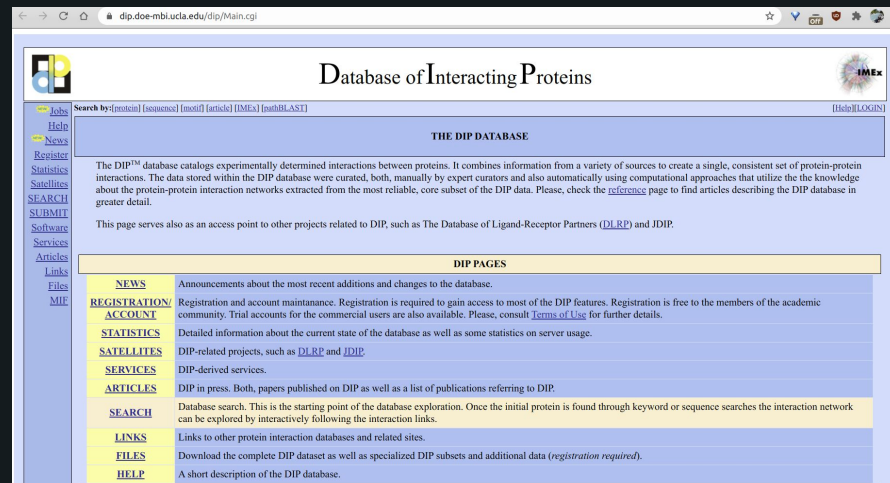
- Bioinformatics Resource Center Impact Survey [Survey](#)
- 2021: Respiratory Pathogen Webinar Series [Webinar](#)
- CDC Flu News & Spotlights
- WHO Global Influenza Surveillance and

# Tools: Examples by field of interest

- Tools by fields of interest:
  - Virology = Microbiology → 65



The screenshot shows the SILVA database homepage. At the top, there are logos for SILVA, Elixir, and de.NBI. The main content area is divided into sections: 'Welcome to the SILVA rRNA database project', 'SILVA provides comprehensive, quality checked and regularly updated datasets...', 'SILVA are the official databases of the software package ARB.', and 'SILVA Alignment, Classification and Tree (ACT) Service'. There is also a 'News' section on the right with recent updates and a 'Merry Christmas & Healthy New Year 2021' message.



The screenshot shows the Database of Interacting Proteins (DIP) homepage. At the top, there is a search bar and a navigation menu. The main content area is titled 'THE DIP DATABASE' and contains a description of the database, its purpose, and a list of related projects. There is also a 'DIP PAGES' section with links to various resources.

# Tools: *Examples by field of interest*


- Tools by fields of interest:
  - Biomedicine → 21

The screenshot shows the PharmGKB website homepage. At the top, a blue banner contains a message about the team's absence during Memorial Day. Below this is a light purple banner with a link to a short tour. The main navigation bar includes the PharmGKB logo and links for Publications, News, Downloads, Contact, and Help. A search bar is prominently displayed with the placeholder text 'Search PharmGKB' and a magnifying glass icon. Below the search bar, a prompt asks the user to search for a molecule, gene, variant, or combination. A Creative Commons license notice is visible, stating that the data is under a CC license and providing links to the Data Usage Policy and citation instructions. At the bottom, four categories are listed with their respective counts: Drug Label Annotations (784), Clinical Guideline Annotations (165), Curated Pathways (153), and Annotated Drugs (715). Each category is accompanied by a small icon representing its content.

← → ↻ 🏠 📄 pharmgkb.org


The PharmGKB team will be out during the week of May 30, returning June 7. We wish all a respectful and safe Memorial Day. ✕





New to the site? [Take a short tour](#) of the home page. ✕

 **Publications** **News** **Downloads** **Contact** **Help**

Search PharmGKB 🔍

Search for a molecule, gene, variant, or combination

 PharmGKB data are under a Creative Commons license. More details are in our [Data Usage Policy](#). Please [cite PharmGKB](#) if you use our information or images.

Drug Label Annotations  <b>784</b>	Clinical Guideline Annotations  <b>165</b>	Curated Pathways  <b>153</b>	Annotated Drugs  <b>715</b>
---	---	---	--

# Tools: *Examples by field of interest*

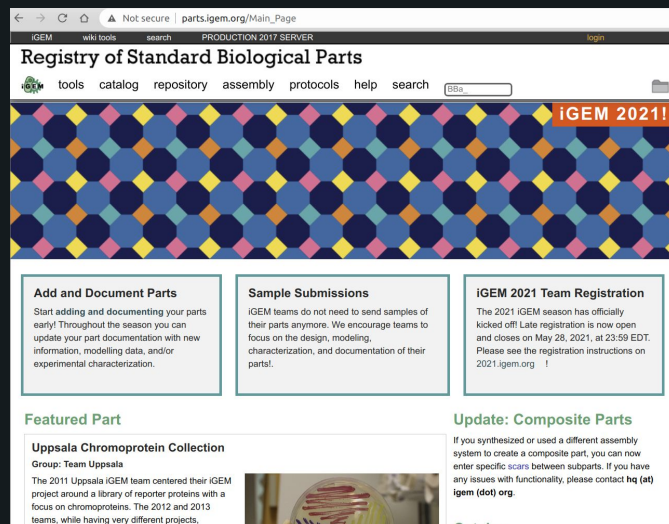
- **Tools by fields of interest:**
  - Synthetic Biology = Genetic Engineering + Systems Biology → 187



The screenshot shows the Cytoscape website homepage. The browser address bar displays 'cytoscape.org'. The navigation menu includes 'Intro', 'Download', 'Apps', 'Documentation', 'Community', 'Report a Bug', and 'Help'. A search bar is labeled 'ENHANCED BY Goo'. The main content area features the Cytoscape logo, the text 'Cytoscape', and the tagline 'Network Data Integration, Analysis, and Visualization in a Box'. There are two prominent buttons: 'Introduction' and 'Download 3.8.2'. A white box contains an 'Important Note for Mac Users' regarding a recent update by Apple causing crashes, with a link to instructions. The background is a complex network graph with various nodes and edges.

Cytoscape is an open source software platform for visualizing complex networks and integrating these with any type of attribute

Welcome Letter



The screenshot shows the iGEM Registry of Standard Biological Parts website. The browser address bar displays 'parts.igem.org/Main\_Page'. The navigation menu includes 'wiki tools', 'search', and 'PRODUCTION 2017 SERVER'. The main content area features a colorful geometric pattern and the text 'Registry of Standard Biological Parts'. There are several sections: 'Add and Document Parts', 'Sample Submissions', 'iGEM 2021 Team Registration', 'Featured Part' (Uppsala Chromoprotein Collection), and 'Update: Composite Parts'. The 'iGEM 2021!' banner is visible in the top right corner.

IGEM 2021!

Add and Document Parts

Start adding and documenting your parts early! Throughout the season you can update your part documentation with new information, modelling data, and/or experimental characterization.

Sample Submissions

iGEM teams do not need to send samples of their parts anymore. We encourage teams to focus on the design, modeling, characterization, and documentation of their parts!

iGEM 2021 Team Registration

The 2021 iGEM season has officially kicked off! Late registration is now open and closes on May 28, 2021, at 23:59 EDT. Please see the registration instructions on [2021.igem.org/](https://2021.igem.org/)

Featured Part

Uppsala Chromoprotein Collection

Group: Team Uppsala

The 2011 Uppsala iGEM team centered their iGEM project around a library of reporter proteins with a focus on chromoproteins. The 2012 and 2013 teams, while having very different projects,

Update: Composite Parts

If you synthesized or used a different assembly system to create a composite part, you can now enter specific scars between subparts. If you have any issues with functionality, please contact [hq@igem.org](mailto:hq@igem.org)

# Tools Catalog: *Featured Insights*

- The universe of bioinformatics tools (which includes computational biology for these purposes) is constantly changing, integrating new tools of varying complexity and level of specialization, this in parallel with the development of its own sub-disciplines, as well as the growing digitization of scientific work in general.
- The tools contained within the catalog comprise mostly **data analysis tools (157/302)**, then general and specialized **database platforms (85/302)**, and thirdly **visualization tools (25/302)**. Then there are a couple of categories that include tools for the **management of internal tasks within laboratories and research teams (15/302)** that are of great importance for the traceability and internal recording of activities (such as Electronic Lab Notebooks, among others). And finally, some tools for **data storage** were included (**6**), **communication tools for the community (4)** (such as ResearchGate), other tools for the purchase of products and **marketplace platforms for life sciences (3)**, and **others (20)**.
- The vast majority of **data analysis and visualization tools are part of a broader process or "pipeline"** within the investigation and study of a particular phenomenon, fulfilling very particular roles, interchangeable with other tools according to the particular needs of a study or the nature of the biological phenomenon studied.
- There are notable generational differences as of 2010, such as better usability for web tools, and open-source libraries available mostly through established repositories such as Github, Gitlab or Bitbucket.

# Tools Catalog: *Featured Insights*

- The countries involved in the generation of a greater part of these tools correspond in descending order to:
  - **The US involved in 48% of the 302 tools, the United Kingdom in 16%, and Switzerland in 6%.**
  - A large part of these tools are linked to institutions such as the NCBI, Broad Institute, EMBL-EBI and the Swiss Institute of Bioinformatics (SIB), which provide financing for their development and/or that host and maintain them within their clusters and digital platforms, helping to give continuity to this work.
- Regarding costs, there is a dominant **majority of tools that are free access or that allow free access for academic purposes: 278/302.**
- Trade-off between free access tools v/s paid tools:
  - Dedicated support and with the possibility of customization to the needs of use for paid tools.
  - **Less sustainability over time for open access tools:**
    - Updating of the tools is fragile, it is given by the availability and interest of its creators.
    - Web hosting is a limitation, some web tools are lost due to lack of funds to give them continuity. Successful cases are given by the availability of their own institutions or other external ones to give access to their servers (such as the EMBL, SIB, NCBI, universities or centers).



## Third specific objective

- Identify the availability of resources to learn how to use these tools and databases, and associated costs.

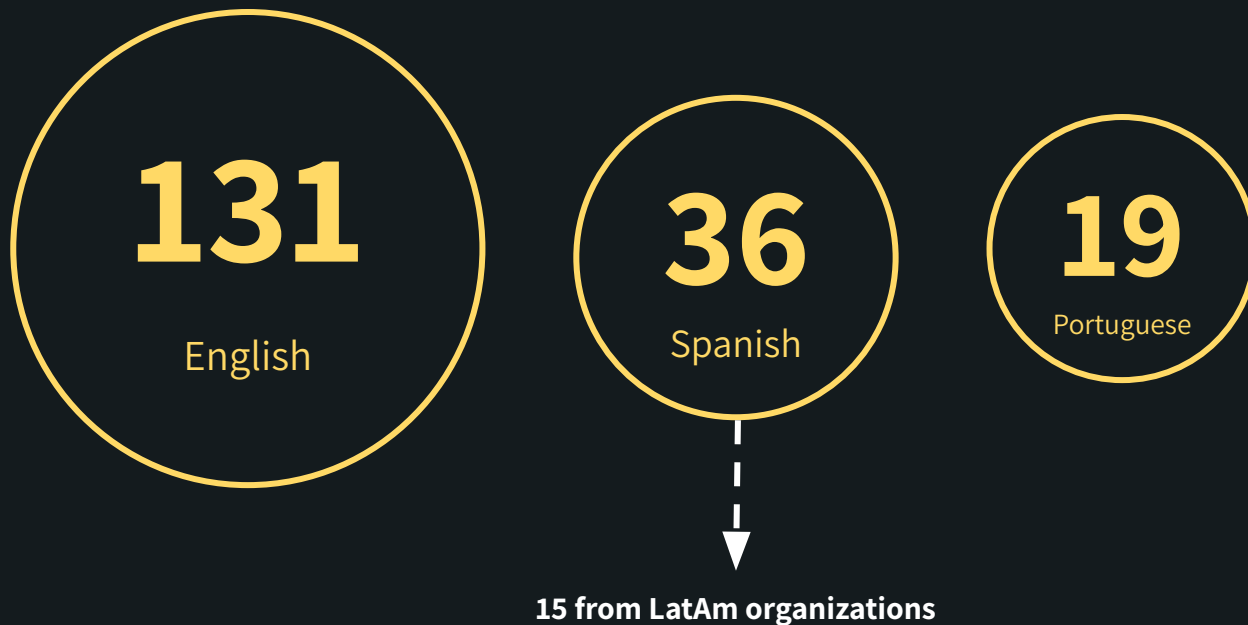
# Learning Resources Catalog

Bioinformatics Toolkit								
Bioinformatics Toolkit v1.0			Bioinformatics Toolkit v2.0			Bioinformatics Toolkit v3.0		
Learning Resources			SHARE			AUTOMATIONS		
VIEWS			By Resource Type (shared)			1 hidden field		
Filter			Group			Sorted by 2 fields		
Color			Share view					
Find a view	Resource Name	Institution/Platform	Resource Type	Modality	Availability	Language	Level	
All	1 Applied Statistical Modelling & Health Informatics MSC, PGCert, ...	King's College London	Diploma	Presential	Timed (yearly/mo...	English	Introductory to the	Advanced training
By Resource...	2 Online Graduate Certificate in Applied Bioinformatics	University of Delaware (UD)	Diploma	Online	Timed (yearly/mo...	English	Introductory to the	Advanced training
Learning Resou...	3 *Graduate Certificate in Bioinformatics (BINF-CERT)	University of Delaware (UD)	Diploma	Presential	Timed (yearly/mo...	English	Introductory to the	Advanced training
(Internal use)	4 *Advanced Certificate in Bioinformatics	Drexel University	Diploma	Presential	Timed (yearly/mo...	English	Introductory to the	Advanced training
Create a view	5 *Certificate In Bioinformatics & Biomarkers	University of Toledo	Diploma		Timed (yearly/mo...	English	Introductory to the	Advanced training
Grid	6 Diplomado en Bioinformática: Aplicaciones e Introducción a las ...	Universidad Manuela Beltrán, Colombia	Diploma	Presential	Timed (yearly/mo...	Español	Advanced training	
Form	7 Diploma de Postitulo en Bioinformática y Biología ...	Universidad de Chile	Diploma	Online	Timed (yearly/mo...	Español	Advanced training	
Calendar	8 Diplomado en Genómica Computacional y Bioinformática	Fundación Arturo Rosenblueth, México	Diploma	Online	On demand	Español	Advanced training	
Gallery	9 Diplomado en Bioinformática	Universidad Nacional Abierta y a Distancia UNA...	Diploma	Online	Timed (yearly/mo...	Español	Advanced training	
Kanban	10 Diplomado En Bioinformática	Universidad Autónoma de Sinaloa. México	Diploma	Presential	Timed (yearly/mo...	Español	Advanced training	
Ga...								
Section								

# Learning Resources Catalog: *Resource type*



# Learning Resources Catalog: *Language*



# Learning Resources Catalog: *Associated costs*

**60**<sub>/149</sub>

With free options: Free +  
Freemium

**73**

Payed resources

Range in USD  
\$12-\$14.839

**16**

Undetermined  
prices

# Learning Resources Catalog: *Featured Insights*

- **Most of the tools for data analysis and visualization coming from the first catalog are part of a process or “pipeline”** within the study of a particular phenomenon, and they fulfill very particular roles in this process, sometimes interchangeable with other tools according to the particular needs of a study or to the nature of the biological phenomenon studied.
  - This reality motivated the need to generate a separate repository for learning resources since many of them must be taught together.
  - This catalog does not include the tutorials or specific user manuals that many of the tools have, they were included within the tools repository in the category “Documentation & Tutorials”, with some exceptions for widely used tools such as R and Python.
- Through the availability of learning resources, **is possible to observe consolidated topics in the field that have extensive learning resources, and emerging topics that have a smaller offer of learning resources:**
  - Consolidated topics: statistics, genomics and analysis of sequencing data in general.
  - Emerging topics: metagenomics, systems biology, molecular simulations, search for new therapeutic targets or drugs, use of artificial intelligence applied to biomedicine.

# Insights about Latin America (LA): *From both catalogs*

- In general, the region is underrepresented both in the development of bioinformatics tools and in the generation of quality learning resources focused on the needs of local scientists.
- Regarding bioinformatics tools:
  - **Authors from LA are involved in the development of only 2.5% of the tools in the catalog, with some cases of international collaborations and no relevant case of intra-regional collaborations.**
  - Most of the catalog tools developed by the region are not widely used by the global community, and they resolve minor technology gaps or for specific use cases.
- Regarding the learning resources available:
  - **A 24% (36/149) of the resources have options in Spanish, of which only 15 resources (including diplomas and courses) have been generated from LA itself.**
  - There are new and interesting cases such as iGEM teams that in order to raise funds are generating courses for the community, and organizations such as Bioscience App and Allbiotech that are also facilitating quality courses in spanish for the LA region.
  - There is a low availability of diploma programs, and local diploma programs address broad topics in a short time (1 year in part-time mode), with a majority focused on the analysis of genomic data, a smaller group on structural biology, and one particular case that includes systems biology.
  - The graduate diplomas are concentrated in Mexico, Colombia and Chile.

# Fourth specific objective

- **Identify skills necessary for the development of new bioinformatics applications, and the adoption of bioinformatics applications developed by others.**
  - Focus group with bioinformatic professionals from the RSG-ISCB network of LatAm:  
June 7, 2021





# Insights from the Focus Group: *LatAm RSG-ISCB Network* \*

- **Total number of participants: 11**
- **Countries represented: Chile, Peru, Colombia, Argentina, Mexico, Venezuela, and Ecuador.**
- **Regarding the digital and computer skills of life science scientists in general (non-bioinformatics), and the adoption of bioinformatics tools:**
  - Experimental scientists generally show a **deficiency in the use of digital tools**. This lack of digital skills includes those like handling Linux systems, console commands, or programming in languages like R or Python.
  - There is a **generational gap and also a language barrier with English**, which is the dominant language, to be able to use these tools and access information such as courses and tutorials.
  - There is **mistrust in the potential and credibility of studies carried out purely *in silico***, which makes their adoption difficult.
  - Undergraduate bioinformatics courses fail to deliver basic skills such as programming, in general they only allow a first approach to theoretical or introductory uses of bioinformatics tools.
  - Proposal: Retrain or re-educate principal investigators who lead the teams and who are the ones who make a large part of the decisions regarding personnel hiring and the research methodologies to be used under their tutelage.

# Insights from the Focus Group: *LatAm RSG-ISCB Network* \*

- **Regarding the job market conditions for LatAm bioinformatic professionals and their training opportunities in the region:**
  - There is a lack of placings for bioinformaticians and research projects that make use of these tools. This is shared by the representatives of Peru, Colombia, Venezuela, Ecuador and Argentina, with the exception of Chile where it is perceived that there is a demand for bioinformatics professionals.
  - Systems for awarding scholarships, allocating funds, thesis project requirements, and others sometimes limit the use of bioinformatics as the primary method and / or the hiring of these professionals.
  - There has been a growth in the number of professionals who are trained in bioinformatics, and of new training programs within the region, however in many countries there are not enough job opportunities for all those who are trained.
  - Brazil, Mexico and Chile are the countries considered leaders in the region in research on the subject.
  - Argentina and Chile in particular, are pioneers in their training programs dedicated 100% to bioinformatics.
  - The lack of renewal of positions limits the entry of young professionals trained in these disciplines and of professionals with greater knowledge of their potential.
  - Bioinformaticians insertion in the private sector is still limited.
  - Bioinformaticians hired in some cases make up for other lacks in infrastructure and access to technical service at universities and centers that do not correspond to their positions.

# Insights from the Focus Group: *LatAm RSG-ISCB Network* \*

- **Regarding the technical infrastructure necessary to make use of bioinformatics tools in the different fields and at different levels of complexity:**
  - The necessary infrastructure is complex and expensive to maintain, which is not always visible to decision makers. Some limiting factors include:
    - Infrastructure for the storage of large amounts of data, and their corresponding backups.
    - Personnel in charge of maintenance.
    - Large quantity and quality of processors required for demanding analysis and complex studies.
    - Security systems to prevent equipment from being damaged.
    - Stable internet connection: Internet access is limited in certain countries, particularly outside capital cities, limiting access to tools and the use of local clusters from other regions.
    - Access to financing from public funds for computers, hard drives, and other electronics.
  - Investments have been made in clusters for academic use in Peru, Colombia and Ecuador, however they have not yielded the expected results due to factors such as the ones stated above, and due to the lack of knowledge/awareness of the community and their own institutions on how to get the most out of these facilities.
  - The development of tools is still led by developed countries because it requires resources such as: high-power clusters and research funds that give researchers the freedom to create scientific results in the form of tools or information for the improvement of other analyzes, and not for necessarily for publication in scientific journals.
  - Proposal: Partner within the region to generate regional clusters that everyone can access through the internet.

# Insights from the Focus Group: *LatAm RSG-ISCB Network* \*

- **Regarding the access to these tools and to training resources for the use of them:**
  - Very broad courses in bioinformatics in general tend to be superficial due to the broad scope of the field and they also fail to generate the necessary skills.
  - Available tutorials are mostly in English and cryptic as they do not explain the developed process in detail and assume the management of certain skills.
  - Proposals:
    - Teach programming as part of the base curriculum for scientific careers.
    - Include R in statistics course syllabi as a first step.
    - Create bootcamps in bioinformatics that in a period of only 6 months can deliver enough skills to professionals in the scientific area to be employable.

*\*All the “insights” shared here were part of the conversation carried out with bioinformatics professionals from Latin America and the Caribbean, belonging to the RSG-ISCB network. Thus, the information shared in this section reflects only the opinion of its authors based on their personal experiences in academia in their respective countries, and not a judgment exercised by those in charge of this study or by the IDB.*

# *Final comments:*

*\*\* The information and statistics provided in this presentation correspond to the study completed in June 2021.*

*\*\*\* The information in both catalogs, made available through the Inter-American Development Bank platform, was updated between September and October 2021, including new tools based on user suggestions.*



**IDB**

Inter-American  
Development Bank