Supplementary File S1

User research (Google Analytics, year 2018)

Users

Parameters	**Numbers**
Total number	531,324
Session/user	~2
Session duration	~3 minutes
Bounce rate	~50% (very good)

Geography

Country	**% Users**
United States	22%
China	11%
India	10%
European countries	17%

Devices

Type	**% Users**
Desktop	94%
Mobile	5%
Tablet	1%

Acquisition

Type	**% Users**
Organic search	56%
Direct	32%
Referral	21%, coming from: • 61% web.expasy.org • 5% prosite.expasy.org • 3% enzyme.expasy.org • 2% NCBI

Behavior flow (all pages)

Pages	% Users
/	35%
/<category>	20%, among which: • 80% /proteomics • 13% /genomics • 4% /transcriptomics
/<category>/<sub-category>	13%, among which: • 22% /proteomics/protein_structure • 20% /proteomics/protein_sequences_and_identification • 13% /genomics/sequence_alignment • 8% /proteomics/protein_characterisation_and_function • 6% /proteomics/post-translational_modification • 6% /proteomics/families__patterns_and_profiles • 6% /proteomics/similarity_search_alignment • 4%/proteomics/protein-protein_interaction

Behavior flow (source)

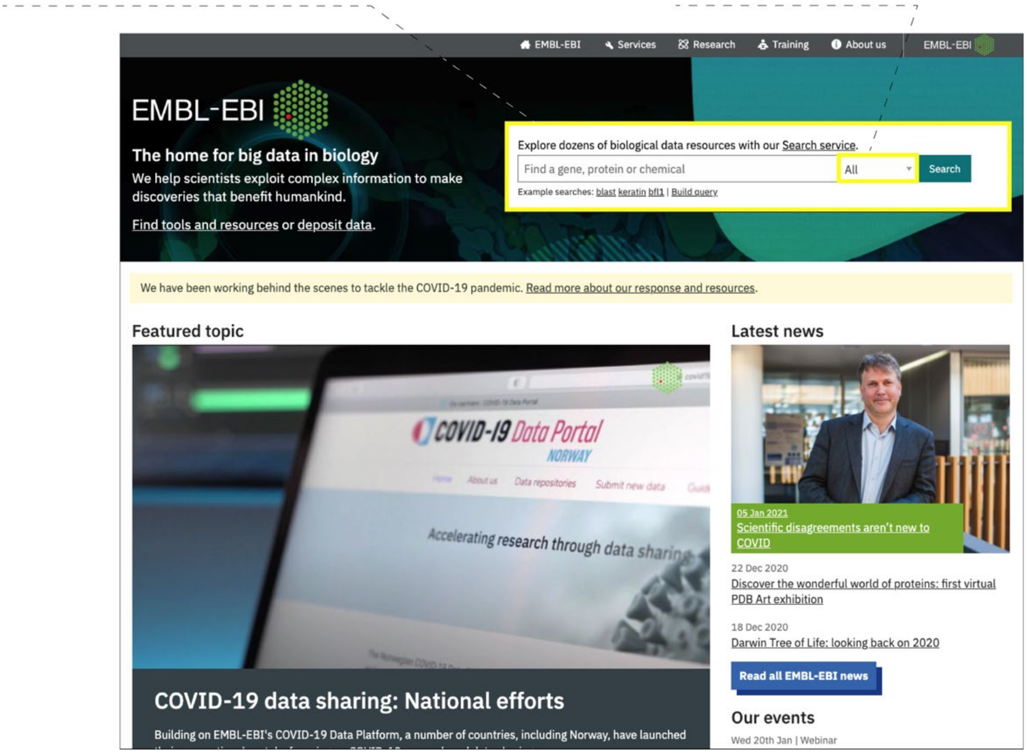
Source type	
Organic search	• 54% land on Home page • 5% land on /resources/search (Home page, "Find resource" selected) • 5% land on /proteomics/protein_structure • 27% land on 100+ different pages.
Direct access	• 64% to the home page • 12% to the /tools page • 6% to the /proteomics page

Supplementary File S2

Expasy's Competitor Benchmarking

EMBL-EBI (https://www.ebi.ac.uk/)
The mission of EMBL-EBI, is to (i) make the world's public biological data freely available to the scientific community via a range of services and tools, (ii) perform basic research and (iii) provide professional training in bioinformatics. The EMBL-EBI website allows both the search for information on the institute itself (i.e. career pages, news, directory, ...) as well as the cross-resources search, and the discovery of the resources developed within the EMBL-EBI. A search bar is located at the top of the page and by default launches the 3 search types described above. A drop-down menu allows the user to refine his search by:

- Scientific search with a list of categories such as Genomic or Small molecule, which launches a cross-resources search,
- EBI content search, which allows, among other things, to discover the resources developed within the EMBL-EBI.



Top-right search bar allowing 3 types of searches:
- General information about EMBL-EBI (career pages, news, people, ...)
- Cross-resources search
- Regular search (explore the resources within the EMBL-EBI)

Dropdown box to refine the search:
- Scientific search (cross-resources search)
- EBI content search (regular search)

The home for big data in biology

We help scientists exploit complex information to make discoveries that benefit humankind.

Find tools and resources or deposit data.

Explore dozens of biological data resources with our Search service.

We have been working behind the scenes to tackle the COVID-19 pandemic. Read more about our response and resources.

COVID-19 data sharing: National efforts

Building on EMBL-EBI's COVID-19 Data Platform, a number of countries, including Norway, have launched

05 Jan 2021
Scientific disagreements aren't new to COVID

22 Dec 2020
Discover the wonderful world of proteins: first virtual PDB Art exhibition

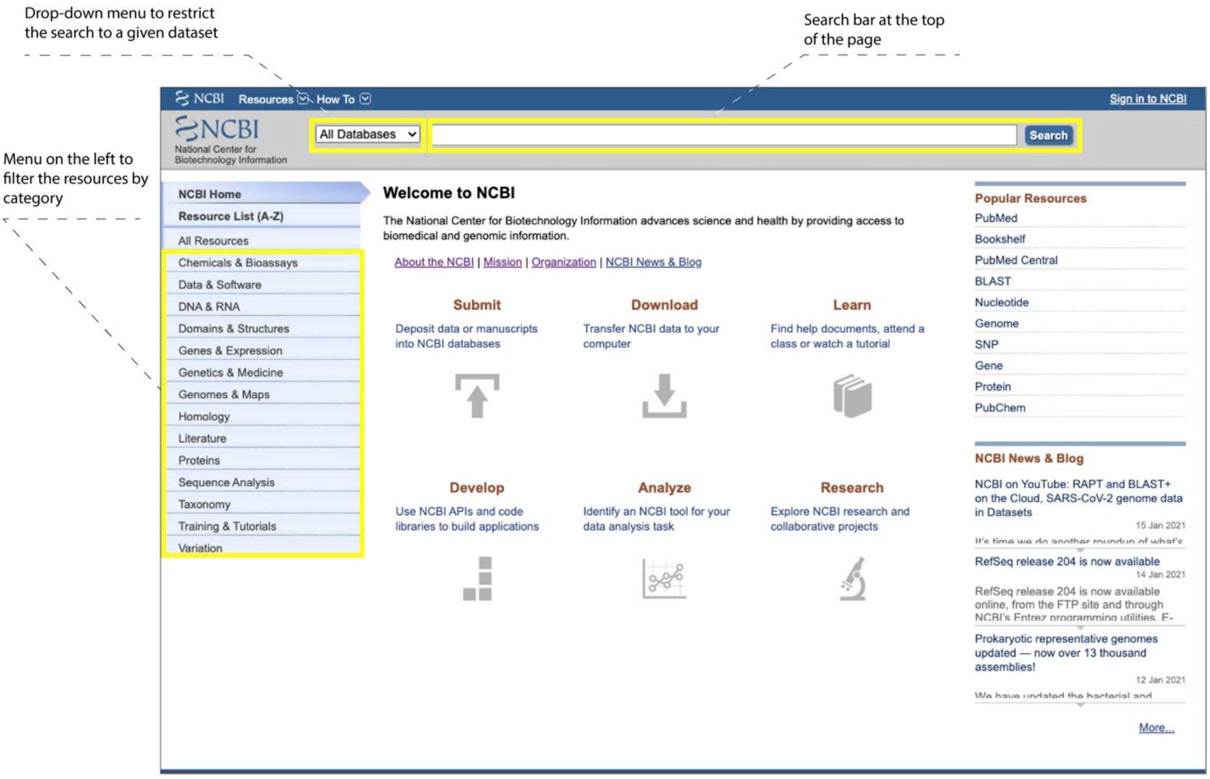
18 Dec 2020
Darwin Tree of Life: looking back on 2020

Wed 20th Jan | Webinar

Figure 1: EMBL-EBI home page

NCBI (https://www.ncbi.nlm.nih.gov/)
NCBI's mission is, among other things, to create automated systems for the storage and analysis of knowledge in molecular biology, biochemistry and genetics and to facilitate the use of these databases and software by the research and medical community. The NCBI website includes a search bar at the top of the page. If the user wishes to limit their search to a certain dataset, they can select it from a drop-down menu. By default, the search performed is of type "cross-resources search". But in case a resource name (such as "BLAST") is detected, the corresponding web resource is described via a card at the top of

the results page. If a user is interested in exploring the resources developed within the NCBI framework, a menu on the left allows to filter the resources by category. The home page also provides a list of popular resources as well as some NCBI news.



Drop-down menu to restrict the search to a given dataset

Search bar at the top of the page

Menu on the left to filter the resources by category

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

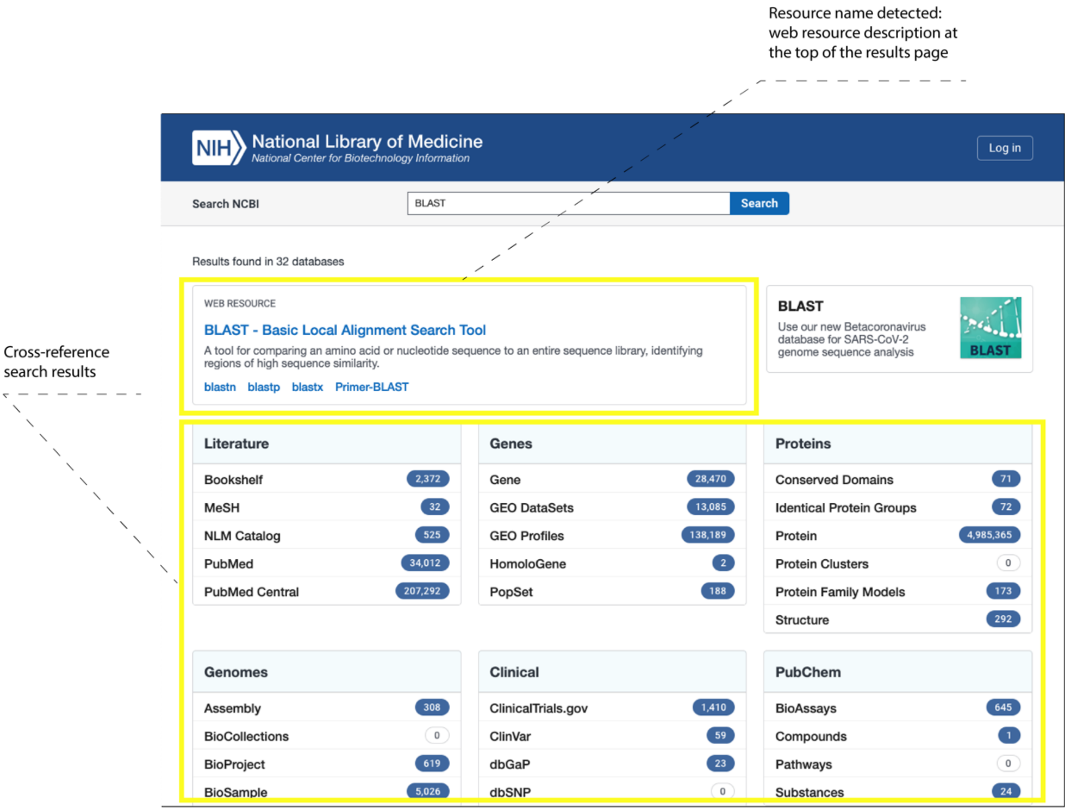
NCBI News & Blog

NCBI on YouTube: RAPT and BLAST+ on the Cloud, SARS-CoV-2 genome data in Datasets
15 Jan 2021
It's time we do another roundup of what's

RefSeq release 204 is now available
14 Jan 2021
RefSeq release 204 is now available online, from the FTP site and through NCBI's Entrez programming utilities, E-

Prokaryotic representative genomes updated — now over 13 thousand assemblies!
12 Jan 2021
We have updated the bacterial and

Figure 2: NCBI home page



Resource name detected: web resource description at the top of the results page

Cross-reference search results

Results found in 32 databases

WEB RESOURCE
BLAST - Basic Local Alignment Search Tool
A tool for comparing an amino acid or nucleotide sequence to an entire sequence library, identifying regions of high similarity.
blastn blastp blastx Primer-BLAST

BLAST
Use our new Betacoronavirus database for SARS-CoV-2 genome sequence analysis

Literature	
Bookshelf	2,372
MeSH	32
NLM Catalog	525
PubMed	34,012
PubMed Central	207,292

Genes	
Gene	28,470
GEO DataSets	13,085
GEO Profiles	138,189
HomoloGene	2
PopSet	188

Proteins	
Conserved Domains	71
Identical Protein Groups	72
Protein	4,985,365
Protein Clusters	0
Protein Family Models	173
Structure	292

Genomes	
Assembly	308
BioCollections	0
BioProject	619
BioSample	5,026

Clinical	
ClinicalTrials.gov	1,410
ClinVar	59
dbGaP	23
dbSNP	0

PubChem	
BioAssays	645
Compounds	1
Pathways	0
Substances	24

Figure 3: NCBI result page

Bio.tools (https://bio.tools/)
The bio.tools portal provides a comprehensive registry of software and databases, facilitating researchers from across the spectrum of biological and biomedical science to find, understand, utilise and cite the resources they need in their day-to-day work. A search bar allows the user to search for all information related to the resources. Through autocompletion, the user can limit the search to a certain type of information (such as Name, Topic, Operation, Input, Output, to name a few). In bio.tools, the resources are described using EDAM terms. All terms are clickable in the resource entries, allowing the user to find more resources with a given term. The resource curation is community driven and is based on more than 1000 contributors. As of January 2021, 16570 resources were described in bio.tools.

Use of EDAM terms for the resource descriptions

What is bio.tools?

The use of bioinformatics is ubiquitous within the life sciences. In *bio.tools*, we are striving to provide a **comprehensive registry** of software and databases, facilitating researchers from across the spectrum of biological and biomedical science to **find**, **understand**, **utilise** and **cite** the resources they need in their day-to-day work.

Everything from simple command-line tools and online services, through to databases and complex, multi-functional analysis workflows is included. Resources are described in a rigorous semantics and syntax, providing end-users with the convenience of **concise**, **consistent** and therefore **comparable** information.

Each *bio.tools* entry is assigned a human-readable, **unique identifier** based on the resource name, *e.g.* biotools:signalp. These identifiers provide a persistent reference to our **"Tool Cards"** of essential information, as well as a means to trace resources and integrate *bio.tools* data with other resources.

All the *bio.tools* data and technical components are available under **open license** and we warmly welcome you to get involved. *bio.tools* development is supported by ELIXIR - the European infrastructure for life science information.

bio.tools is supported by ELIXIR through the ELIXIR-EXCELERATE grant, which is funded by the European Union Horizon 2020 program under grant agreement 676559.

Open Data

bio.tools content is freely available to all under CC BY 4.0 license - you are free to share and adapt the data, so long as you give credit and don't restrict the freedom of others.

Open Source

bio.tools source code is freely available to all under GPL-3.0 - you are free to share and adapt our software, but you must ensure it remains free for all its users.

Built by You

We depend on the goodwill and enthusiasm of our 1000+ (and growing!) contributors - if you develop or provide tools and online services, please add them after signing-up.

Tool IDs

All *bio.tools* entries are assigned a human-friendly unique identifier, *e.g.* biotools:signalp. Once verified, a *bio.tools* ID provides a stable way to trace resources and integrate *bio.tools* data with other projects.

Standard Semantics

The scientific function of *bio.tools* resources can be precisely annotated in defined terms from the EDAM ontology, including common topics, operations, types of data and data formats.

Standard Syntax

bio.tools resource descriptions adhere to a rigorous syntax defined by biotoolsSchema, which provides regular expressions, controlled vocabularies and other syntax rules for 50 key attributes.

Community-driven

We rely upon scientific communities to improve the terminology and description of resources in different domains of the life sciences - we welcome your help with this work in progress.

Backed by ELIXIR

bio.tools is anchored within ELIXIR, the European Infrastructure for Biological Information. *bio.tools* will remain free, open and maintained in the long term.

Tools Platform

bio.tools is an integral part of the ELIXIR Tools Platform, enabling the development, description, discovery, re-use, deployment and benchmarking of software tools and workflows.

Figure 4: Bio-tools home page

The BioCatalogue (https://www.biocatalogue.org/)
The BioCatalogue is a curated catalogue of life science web services. The BioCatalogue website is freely accessible to the world thanks to a powerful search engine. Expert curators and users join forces to provide high quality annotations for services. A search bar is located at the top of the page, together with a link which leads to a list of all the BioCatalogue web services, which can later on be filtered by category thanks to a menu on the left. As of January 2021, 1431 web services were described in the BioCatalogue.

The BioCatalogue: providing a curated catalogue of life science Web services

BioCatalogue currently has 1431 services, 266 service providers and 911 members

Helpful Links

Getting started with the Catalogue
General info on Web services
Best practice guide for developing Web services
SSI's top tips on creating Web services
Executing Web services
Turn your command line application(s) into Web services

Citing BioCatalogue

Bhagat, J., Tanoh, F., Nzuobontane, E., Laurent, T., Orlowski, J., Roos, M., Wolstencroft, K., Aleksejevs, S., Stevens, R., Pettifer, S., Lopez, R., Goble, C.A.
BioCatalogue: a universal catalogue of web services for the life sciences
Nucl. Acids Res. (2010) 38: 689-694
doi:10.1093/nar/gkq394

"Web Services are hard to find"
DISCOVER
- Find the right Web Service
- Powerful search and filtering
- Information from providers and community

More info

"My Web Services are not visible"
REGISTER
- Easily register Web Services
- Instantly available to everyone
- Providers can advertise, describe and monitor their Services

More info

"Web Services are poorly described"
ANNOTATE
- Anyone can describe and annotate
- Ongoing expert curation
- Social curation by the community

More info

"Web Services are volatile"
MONITOR
- Services change and get outdated
- BioCatalogue monitors Services
- Monitors availability and reliability

More info

Site Announcements

ELIXIR Tools registry 'community paper' published in NAR
By Niall Beard (about 5 years ago)

BiodiversityCatalogue presents development of Web Map Service support at TDWG
By Niall Beard (over 5 years ago)

Scheduled Maintenance
By Niall Beard (almost 7 years ago)

BioCatalogue Rails 3 Upgrade
By Aleksandra Nenadic (about 7 years ago)

Soaplab EMBOSS services end of life: 31st January 2013
By Robert Haines (about 8 years ago)

More

Copyright (c) 2008 - 2021 The University of Manchester and the European Bioinformatics Institute (EMBL-EBI)

The project was funded by the BBSRC (BB/F01046X/1, BB/F010540/1)

Figure 5: Biocatalogue home page

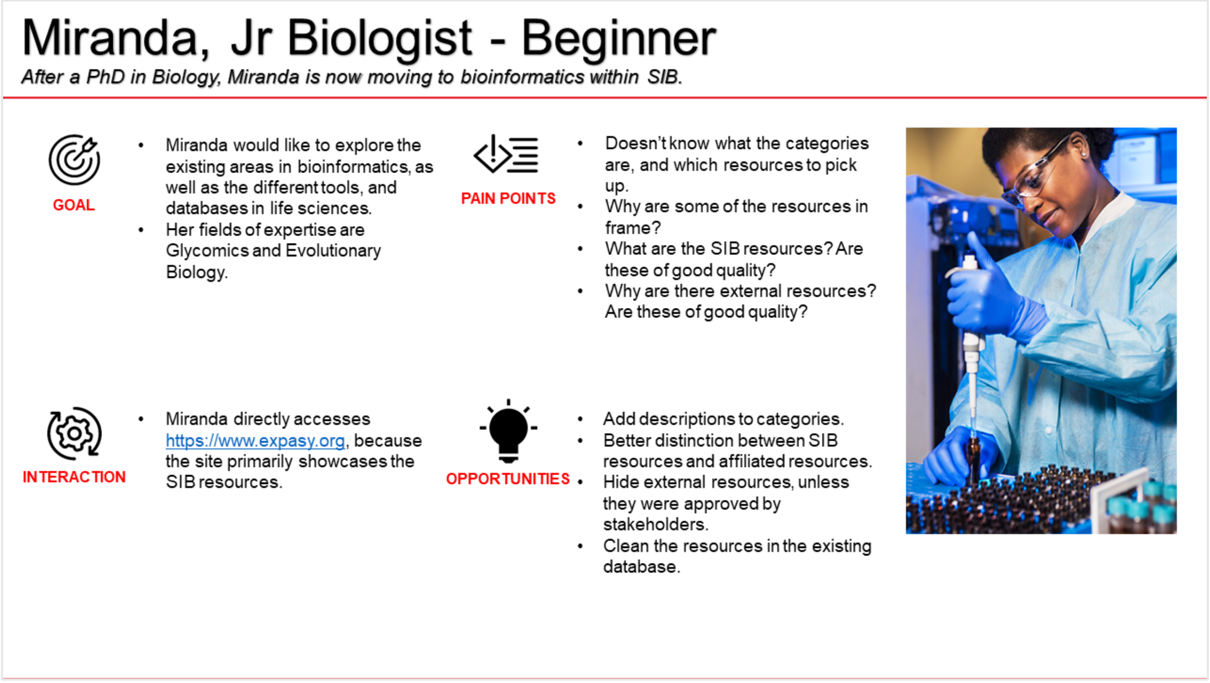
The strengths of each of the four competitors are shown in Table 1

Table 1: Good practices found in the four Expasy's competitors

Website	**Strengths**
EMBL-EBI	Unique search bar for different types of searches (institutional information, cross-reference and regular search)
NCBI	Unique search bar (cross-reference and regular search) Filter by categories on the left Joint results page (cross-reference and regular search)
Bio.tools	Use of EDAM terms for the resource description
Biocatalogue	Expert-curated catalogue

Supplementary File S3

Personas



Miranda, Jr Biologist - Beginner

After a PhD in Biology, Miranda is now moving to bioinformatics within SIB.

GOAL

- Miranda would like to explore the existing areas in bioinformatics, as well as the different tools, and databases in life sciences.
- Her fields of expertise are Glycomics and Evolutionary Biology.

PAIN POINTS

- Doesn't know what the categories are, and which resources to pick up.
- Why are some of the resources in frame?
- What are the SIB resources? Are these of good quality?
- Why are there external resources? Are these of good quality?

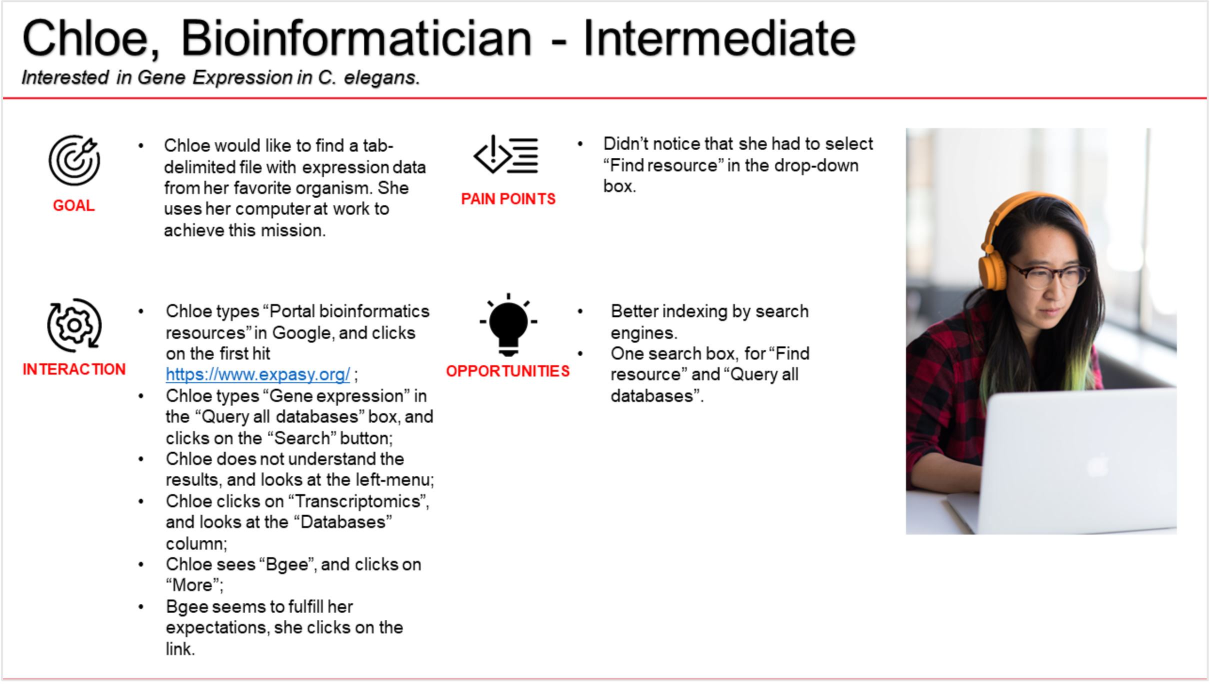
INTERACTION

- Miranda directly accesses https://www.expasy.org, because the site primarily showcases the SIB resources.

OPPORTUNITIES

- Add descriptions to categories.
- Better distinction between SIB resources and affiliated resources.
- Hide external resources, unless they were approved by stakeholders.
- Clean the resources in the existing database.

Figure 1: Miranda, Jr Biologist – Beginner. Photo credits: unsplash.com



Chloe, Bioinformatician - Intermediate

Interested in Gene Expression in C. elegans.

GOAL

- Chloe would like to find a tab-delimited file with expression data from her favorite organism. She uses her computer at work to achieve this mission.

PAIN POINTS

- Didn't notice that she had to select "Find resource" in the drop-down box.

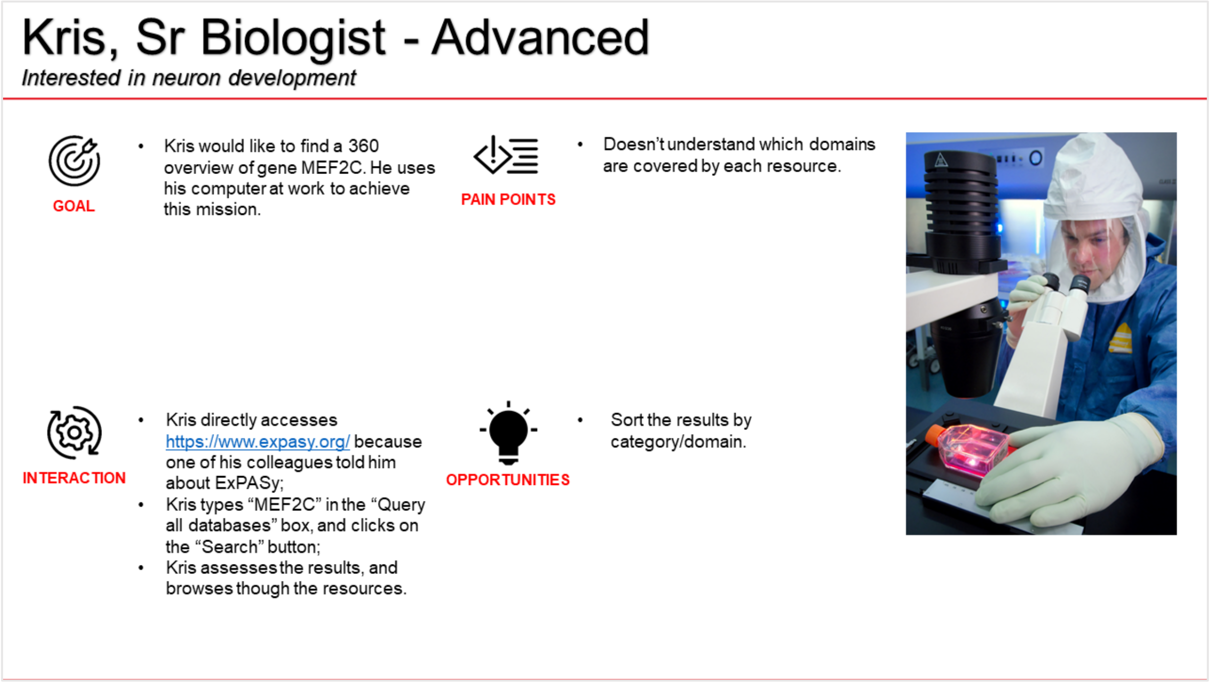
INTERACTION

- Chloe types "Portal bioinformatics resources" in Google, and clicks on the first hit https://www.expasy.org/ ;
- Chloe types "Gene expression" in the "Query all databases" box, and clicks on the "Search" button;
- Chloe does not understand the results, and looks at the left-menu;
- Chloe clicks on "Transcriptomics", and looks at the "Databases" column;
- Chloe sees "Bgee", and clicks on "More";
- Bgee seems to fulfill her expectations, she clicks on the link.

OPPORTUNITIES

- Better indexing by search engines.
- One search box, for "Find resource" and "Query all databases".

Figure 2: Chloe, Bioinformatician - Intermediate Photo credits: unsplash.com



Kris, Sr Biologist - Advanced

Interested in neuron development

GOAL

- Kris would like to find a 360 overview of gene MEF2C. He uses his computer at work to achieve this mission.

PAIN POINTS

- Doesn't understand which domains are covered by each resource.

INTERACTION

- Kris directly accesses https://www.expasy.org/ because one of his colleagues told him about ExPASy;
- Kris types “MEF2C” in the “Query all databases” box, and clicks on the “Search” button;
- Kris assesses the results, and browses though the resources.

OPPORTUNITIES

- Sort the results by category/domain.

Figure 3: Kris, Sr Biologist – Advanced Photo credits: unsplash.com

Table 1: Personas are built based on the insights from the preliminary user research. The table describes how the data from Google Analytics study are reflected in the personas’ characteristics and/or behaviours.

Insights from user research (Google Analytics)	Personas’ characteristics and/or behaviours	
Acquisition: Direct (32%)	**Miranda**	Accesses the website directly
Devices: Desktop (94%)	**Chloe**	Uses a Desktop
Acquisition: Organic search (56%)		Uses Google
User behaviour: 54% of Google users land on the home page		Lands on the home page
User behaviour: 20% of Google users click on a category		Clicks on Transcriptomics
Acquisition: Direct (32%)	**Kris**	Accesses the website directly
User behaviour: 15% of direct accesses use the search functionality		Types a search term in the search box