BIOMATDB

ADVANCED DATABASE FOR BIOMATERIALS WITH DATA ANALYSIS AND VISUALISATION TOOLS EXTENDED BY A MARKETPLACE WITH DIGITAL ADVISORS

**Grant Agreement: 101058779**

# D3.2 Database, Data Processing Methods, Tools and Web Application Marketplace and Specifications (Phase 1)

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

## Project description

| | |
|---|---|
| Acronym | **BIOMATDB** |
| Title | **Advanced Database for Biomaterials with Data Analysis and Visualisation Tools extended by a Marketplace with Digital Advisors** |
| Coordinator | SYNYO GmbH |
| Reference | 101058779 |
| Type | Coordination and Support Action (CSA) |
| Programme | **Horizon Europe (HORIZON)** |
| Topic | HORIZON-CL4-2021-RESILIENCE-01-25<br>Biomaterials Database for Health Applications |
| Start | 01.06.2022 |
| Duration | 30 months |
| Website | www.biomatdb.eu |
| Consortium | **SYNYO GmbH (SYNYO)**, Austria, Coordinator<br>**Barcelona Supercomputing Center** (BSC), Spain<br>**University of Oslo** (UIO), Norway<br>**Universitat Politècnica de Catalunya** (UPC), Spain<br>**Tampere University** (TAU), Finland<br>**West Pomeranian University of Technology**, **Szczecin** (ZUT), Poland<br>**MINDS & SPARKS GmbH** (M&S), Austria<br>**Business Council of the Center Region & CCIC** (CEC/CCIC), Portugal<br>**Clust-ER Health** (CLUSTER), Italy<br>**National University of Ireland**, **Galway** (NUIG), Ireland<br>**Hospital Infantil Universitario Niño Jesús** (FHUNJ), Spain<br>**European Connected Health Alliance** (ECHA), Ireland<br><br>**Osnat Hakimi** (Subcontractor), Israel<br>**Carla Verónica Fuenteslópez** (Subcontractor), United Kingdom<br>**Yannis Missirlis** (Subcontractor), Greece |

## Deliverable

| | |
|---|---|
| Deliverable number | **D3.2** |
| Deliverable title | **Database, Data Processing Methods, Tools and Web Application Marketplace and Specifications (Phase 1)** |
| Deliverable version | 1.0 |
| Lead beneficiary | **BSC** |
| Work package number | WP3 |
| Work package title | CONCEPT: Features, Technical Systems, Advanced Database, Analysis Tools, Marketplaces, and Digital Advisors |
| Due date of delivery | 31.05.2023 |
| Actual date of delivery | 31.05.2023 |
| Dissemination level | Public (PU) |
| Type | **Report** |
| Rights | BIOMATDB Consortium |
| Authors | **Jan Rodríguez** (BSC)<br>**Miguel Rodríguez** (BSC)<br>**Martin Krallinger** (BSC)<br>**Jakob Seper** (SYNYO)<br>**Tilman Kerl** (SYNYO)<br>**Thomas Paulin** (SYNYO) |
| Contributors | **Peter Leitner** (SYNYO)<br>**Tomas Lima** (SYNYO)<br>**Darryl Estrada** (BSC)<br>**Ali Kazemi** (SYNYO)<br>**Dan Prihoi** (SYNYO) |
| Reviewers | **Juan Manuel Uribe-Gomez** (NUIG)<br>**Celine Rabé** (MINDS & SPARKS)<br>**Violeta Heinze** (MINDS & SPARKS)<br>**Damir Haskovic** (MINDS & SPARKS) |

# Executive Summary

This deliverable aims to specify the detailed data structures and formats including all relevant biomaterial properties and entities pertaining to the technical solutions developed within the ambit of the BIOMATDB project – the Biomaterial Database and the Biomaterial Marketplace. It also covers the definition of related analytical methods and suitable visualisation techniques. Concerning the Biomaterial Database, this includes data structures, data formats and entities, processing methods, frameworks and visualisations. For the Biomaterial Marketplace, the deliverable encompasses a comprehensive overview of the technical system and its entities as well as an outline of the corresponding web application, placing significant emphasis on the supportive technologies and components involved.

In terms of the Biomaterial Database (will be accessible via www.biomaterialdatabase.com), advanced Natural Language Processing (NLP) techniques were used to enhance information retrieval, categorisation and semantic indexing of biomaterials-related content. The report highlights key considerations such as data sources, data flow, content selection and pre-processing. Techniques including topic modelling, text similarity, text classification, clustering, named entity recognition (NER) and metadata/entity normalisation were employed, supported by iterative improvements through guidelines, annotations and machine learning models. The report also presents annotation tools, software and the initial design of the database management system. Additionally, it explores frameworks and supporting libraries for development, along with technologies and techniques for interactive visualisation in the web-interface.

Regarding the Biomaterial Marketplace (will be accessible via www.biomaterialmarketplace.com), the technical approach is centred around three overarching meta use cases, the administration panel, the marketplace and the digital advisors. In addition to a detailed description and demonstration of the application modules, the report highlights the significance of data sources, comprising core modules manually inputted and extension modules sourced externally. Subsequently, the progression of the data flow, leading up to the presentation to the end user, is explained. Furthermore, various aspects including the web application's user experience, user interface design, development patterns and special features are covered in this deliverable. It also conducts a thorough analysis of state-of-the-art frameworks and technologies across system-, frontend- and analytics components.

The results of this deliverable and the introduced concepts, tools and processes will be used in the consecutive activities of WP3 (Phase 2), such as design and specification of the backend-components for the Biomaterial Database and the Biomaterial Marketplace. Moreover, the outputs of this deliverable will support activities in WP4, mainly D4.3, the description of the implementation of the frontend interfaces using modern web application frameworks, and D4.4 – the integration of relevant biomaterial data and entities.

# Contents



# Figures



# Tables



# Acronyms & Abbreviations

| Term | Description |
|---|---|
| API | Application Programming Interface |
| ATMP | Advanced Therapy Medicinal Product |
| B2B | Business to Business |
| B2G | Business to Government |
| CORDIS | Community Research and Development Information Service |
| CPV | Common Procurement Vocabulary |
| CSS | Cascading Style Sheets |
| D | Deliverable |
| DDos | Distributed Denial of Service |
| EC | European Commission |
| EU | European Union |
| FAQ | Frequently Asked Questions |
| HTML | Hypertext Markup Language |
| HTTP | Hypertext Transfer Protocol |
| HTTPS | Hypertext Transfer Protocol Secure |
| JSON | JavaScript Object Notation |
| KB | Knowledge Base |
| MP | Marketplace |
| NER | Named Entity Recognition |
| NLP | Natural Language Processing |
| PDF | Portable Document Format |
| SSL | Secure Sockets Layer |
| SVG | Scalable Vector Graphics |
| TED | Tenders Electronic Daily |
| TLS | Transport Layer Security |
| UI | User Interface |
| UX | User Experience |
| WP | Work Package |

# 1 Introduction

## 1.1 Overview

This deliverable primarily focuses on specifying the detailed data structures and formats, including all relevant biomaterial properties and entities, in addition defining related analytical methods and suitable (interactive) visualisation techniques. It covers the Biomaterial Database's data structures, formats, entities, processing methods, frameworks and visualisations, along with the Biomaterial Marketplace's system, main entities and web application. Furthermore, it presents the feature descriptions, input/output data formats, interfaces, process diagrams, and framework selection process for development across all the components.

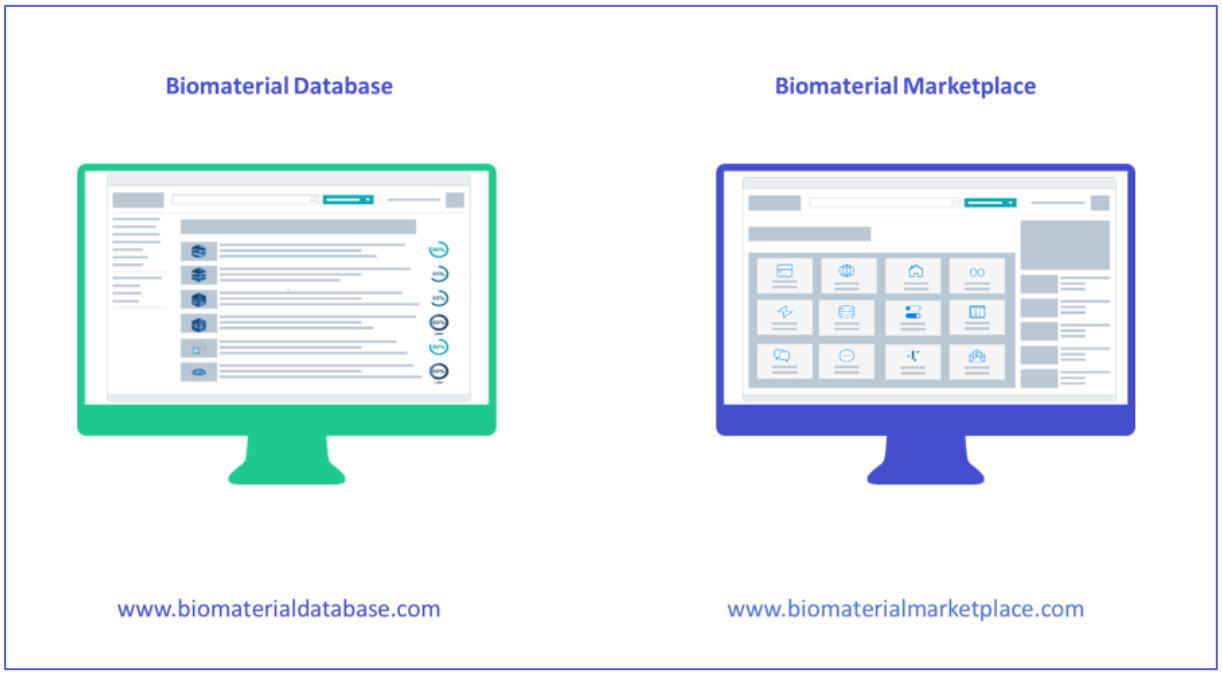



**Figure 1.** Biomaterial Database and Biomaterial Marketplace

The concepts presented in this deliverable are mostly based on current results obtained in (1) WP2, including the definition of requirements, use cases, the labelling approach and the data sources; and (2) WP3, which focuses on the core technical concept of the Biomaterial Database, data analysis tools, as well as the specification and design of the marketplace and digital advisors.

The deliverable will be used as the basis for the next activities in WP3 and WP4. These include the detailed specification of the data structures and formats, as well as the systems and system components. It will also be used in the continuous design of the systems and finally the development of the systems in multiple iterations. Thus, this deliverable will support the subsequent development work in the BIOMATDB project and extends the foundations built so far. This will ensure that the technical infrastructure of the system is well designed and fit to the purpose of the project, while keeping in mind potential issues, such as data security, scalability, and ease of use.

## 1.2 Relation to other tasks and deliverables

This deliverable is related to the following other BIOMATDB tasks and deliverables:

**Receives inputs from:**

**Table 1.** D3.2 Input from other tasks and deliverables

| Deliverable | Due Date | Input for D3.2 |
|---|---|---|
| D2.1 | 31.01.2023 | Knowledge about sources of biomaterials data |
| D2.2 | 31.01.2023 | Requirements of stakeholders and use cases |
| D2.3 | 31.01.2023 | Meta use cases, requirements and label specification report |
| D3.1 | 28.02.2023 | Core technical concept of the Biomaterial Database, data analysis tools, marketplace and digital advisors based on requirements |

**Provides outputs to:**

**Table 2.** D3.2 Output for other tasks and deliverables

| Deliverable | Due Date | Output from D3.2 |
|---|---|---|
| D3.3 | 30.03.2023 | Elaboration of the development tool and framework selection |
| D3.4 | 31.03.2024 | Design of the technical backend-framework (Phase 2) |
| D3.5 | 30.04.2024 | Design of the web-application frontends |

## 1.3 Structure of the deliverable

This deliverable is structured as follows: Chapter 1 outlines the overall purpose and aim of this deliverable. The deliverable is divided into five main chapters that describe the setup and progress of the technical development for the Biomaterial Database, the web-interface to the database, and the biomaterial marketplace (including backend and frontend components).

Chapter 2 focuses on the setup and structure of the development of the Biomaterial Database using advanced NLP techniques. It covers data sources and flow in the development process as well as the selection of relevant content, based on an internal survey, data harvesting methods for sources like PubMed and Clinicaltrials.gov, and the importance of data pre-processing and harmonisation.

Chapter 3 outlines the use of topic modelling, text similarity, text classification, clustering, Named Entity Recognition (NER), and metadata/entity normalisation to enhance information retrieval, categorisation, and semantic indexing of biomaterials-related content. Also, the iterative improvement process involving guidelines, annotations, and machine learning models to refine the generated resources is presented. Moreover, annotation tools, software, and models used in the Biomaterial Database implementation, as well as the initial design of the database management system are outlined. Finally, frameworks and techniques for interactive visualisation for the web-interface to the database are presented and discussed.

Chapter 4 aims to provide an overview of the Biomaterial Marketplace technical approach. This includes various use cases and personas to specify the requirements as well as a holistic outline of the multiple data sources and the subsequent data flow. To finalise the overview of the technical system of the Biomaterial Marketplace, all key entities of the application are described and demonstrated with existing examples.

Chapter 5 deals with all relevant aspects in terms of user experience (UX) of the Biomaterial Marketplace and the web application itself. The chapter covers basic parts of such an application like the user interface (e.g., used templates) but also explains the more sophisticated features of the application, like digital advisors or an intelligent search. Another key part of this chapter is the analysis of technologies and components that can be used to provide the end users the best marketplace application possible. This analysis is divided in three parts: system components, frontend components, and analytics components.

# 2 Biomaterial Database: Use cases, data structures, data formats and entities

Biomaterial-relevant data can be found in different document types and formats, including: txt, JSON, XML, and HTML or PDF. Searching, processing, and analysing each document becomes cumbersome because there are thousands of documents to extract valuable information in terms of biomaterials characteristics. In consequence, robust, fast, and powerful programming language designed for managing and manipulating relational databases such as Structured Query Language (SQL) and non-relational (NoSQL) databases will serve to host unstructured and structured data and content with the aim of building the Biomaterial Database backend.

In order to ensure the saving, updating and retrieving of the data for future use, database systems become crucial. NLP techniques applied on the biomaterials data will provide helpful insight into the type of data we are working with. To achieve this goal, it is essential the use of advanced machine learning text analysis approaches, such as biomaterials concept, entity recognition, text classification and text similarity. To ensure consistency between system design and database design, developing a properly planned database structure that covers a well-defined meta-data schema is important.

The development and core technical concepts of the Biomaterial Database content processing aspects can be summarised in the following tasks:

- **Task 1 (T.1)**: Biomaterials relevant content characterization and selection criteria definition.
- **Task 2 (T.2)**: Data Harvesting.
- **Task 3 (T.3)**: Data pre-processing and harmonisation.
- **Task 4 (T.4)**: Design, implementation and testing of content classification, document triage, text similarity, clustering and ranking components.
- **Task 5 (T.5)**: Design, implementation and testing of automatic semantic content annotation components of biomaterials relevant concept types and entities.
- **Task 6 (T.6)**: Iterative improvement, quality control and adaptation of content classification, document triage and automatic semantic annotation components for biomaterials data.
- **Task 7 (T.7)**: Database backend design, data import, indexing and biomaterials database framework development.
- **Task 8 (T.8)**: Database backend design, web application system development and iterative development of functionalities with user testing and biomaterials expert feedback.
- **Task 9 (T.9)**: Database frontend design, web application system development and iterative development of functionalities with user testing and biomaterials expert feedback.

The underlying database development process is summarised in Figure 2. The different tasks and steps needed to be carried out along the connections between the tasks and their relevance in practice for the meta-use cases supported by the database. Data extraction, harmonisation and NLP techniques to meta use case and front-end functionalities are included among the tasks.

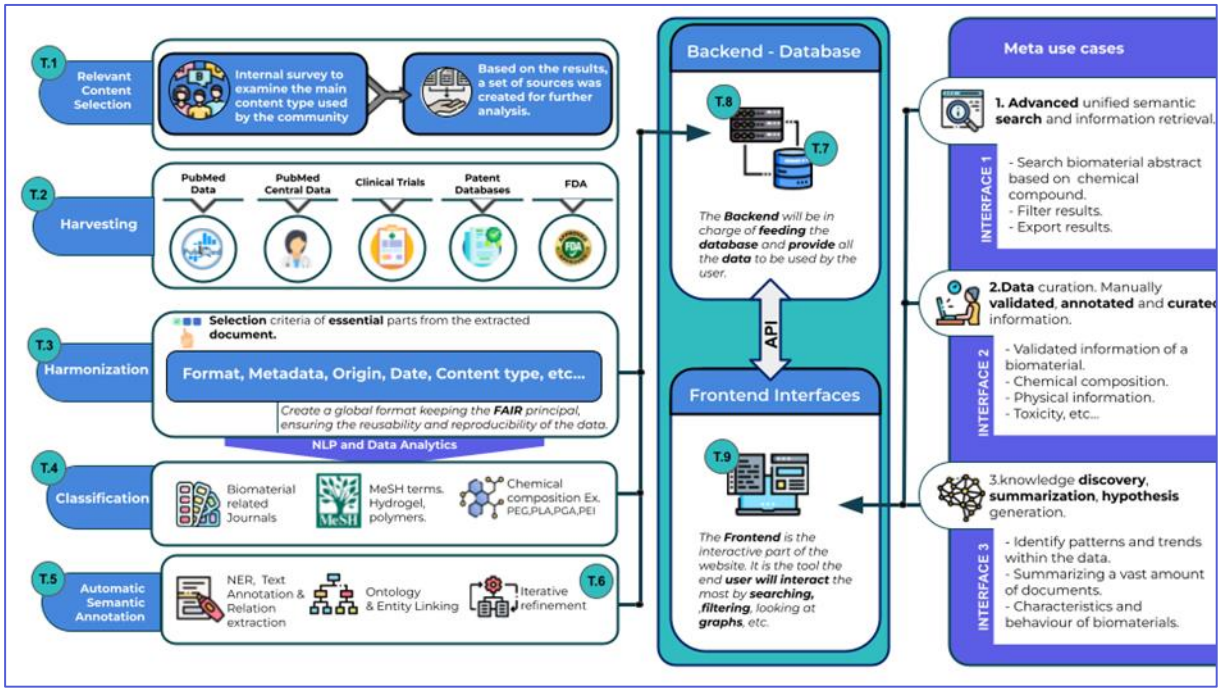



**Figure 2.** Task workflow

The proposed Biomaterial Database development will use automatic content retrieval and data harvesting techniques. Crucially, advanced NPL components, biomedical language models, ontology, and controlled vocabulary will be applied in order to generate a more comprehensive, systematic, and sustainable resource that can cope with the automatic processing of large data collection relevant for biomaterials and their biocompatibility characteristics. Those components will be used to extract and semantically annotate biomaterials-relevant content automatically, to serve as a technical base to enable sophisticated semantic searches, manual curation of biomaterials and biocompatibility information, and enable predictive, data analytics, and knowledge discovery from database functionalities.

## 2.1 Summary of meta use cases

There are three meta use cases defined for the Biomaterial Database, which describe general capabilities that the system will provide. More technical description of the meta use cases is further detailed in the private delivery D2.3.

1. **Advanced Unified Biomaterials Semantic Search and Information Retrieval System**

Research and development of biomaterials and their applications have been increasing dramatically in recent years, resulting in a tremendous volume of published information on the subject. Researchers, on the other hand, may struggle to find relevant and trustworthy information regarding biomaterials and their applications like medical devices. Traditional keyword-based search engines may miss the significance and context of that content, yielding irrelevant or incomplete results. We propose the development of a comprehensive unified biomaterials semantic search and information retrieval system to address this issue. The proposed advanced unified biomaterials semantic search and information retrieval system have the potential to revolutionise the way that researchers access and utilise the information on biomaterials.

The use of natural language processing and machine learning techniques to understand the meaning and context of biomaterials-related content, combined with an intuitive and user-friendly interface, will make it a valuable tool for the biomaterials research community.

2. **Manual Annotation, Validation, and Curation of Biomaterial Information**

It is crucial, in the field of biomaterials, to have accurate and trustworthy information on the materials being utilised or investigated. However, because biomaterials are complex and diverse, it can be difficult to collect and organise this information in a systematic and consistent manner. This meta use case addresses the requirement for biomaterials information that has been carefully annotated, validated, and curated.

The objective of this meta use case is to develop a system for the curation of biomaterials information that is manually annotated, validated, and curated by experts in the field. The system will provide a user-friendly interface for this aim, as well as a database for storing and retrieving the information.

3. **Knowledge discovery, summarization, hypothesis generation**

Biomaterials research is a fast-expanding subject concerned with the investigation of materials that interact with biological systems. Tissue engineering, drug medication delivery, and biomedical imaging are just a few examples of these materials' uses. However, because of the huge volume of data created in this subject, researchers may find challenging to adequately evaluate, interpret and make sense of all the information.

This meta use case is focused on creating a system that leverages advanced machine learning techniques to assist researchers in the biomaterials field with predictive modelling, knowledge discovery, summarization, and hypothesis generation. This system will provide a user-friendly interface that allows easy access to a vast database of biomaterials information, and will allow researchers to quickly and efficiently identify patterns and trends within the data, leading to new insights and discoveries.

## 2.2 Data sources and data flow

Figure 3 schematically illustrates the data flow and sources on a high level. For a more detailed overview or focus on the individual steps and components, please see the respective figures and subsections.

### 2.2.1 Relevant Content Selection

With the aim of prioritising the most relevant content and document sources for building a biomaterials database and content storage, the BIOMATDB consortium has conducted an internal survey to examine which content types are: (1) currently being used and consulted by the biomaterials research community and (2) contain information of importance to characterise relevant biomaterials features, characteristics, and aspects related to biocompatibility evidence and clinical application scenarios. Additionally, results from D2.1. D2.2, and D2.3 were also examined to understand and characterise the main content sources for the biomaterials research community, including concrete data collections and access strategies underlying these datasets. This step refers to Task 1 (T.1) mentioned in the previous introduction.

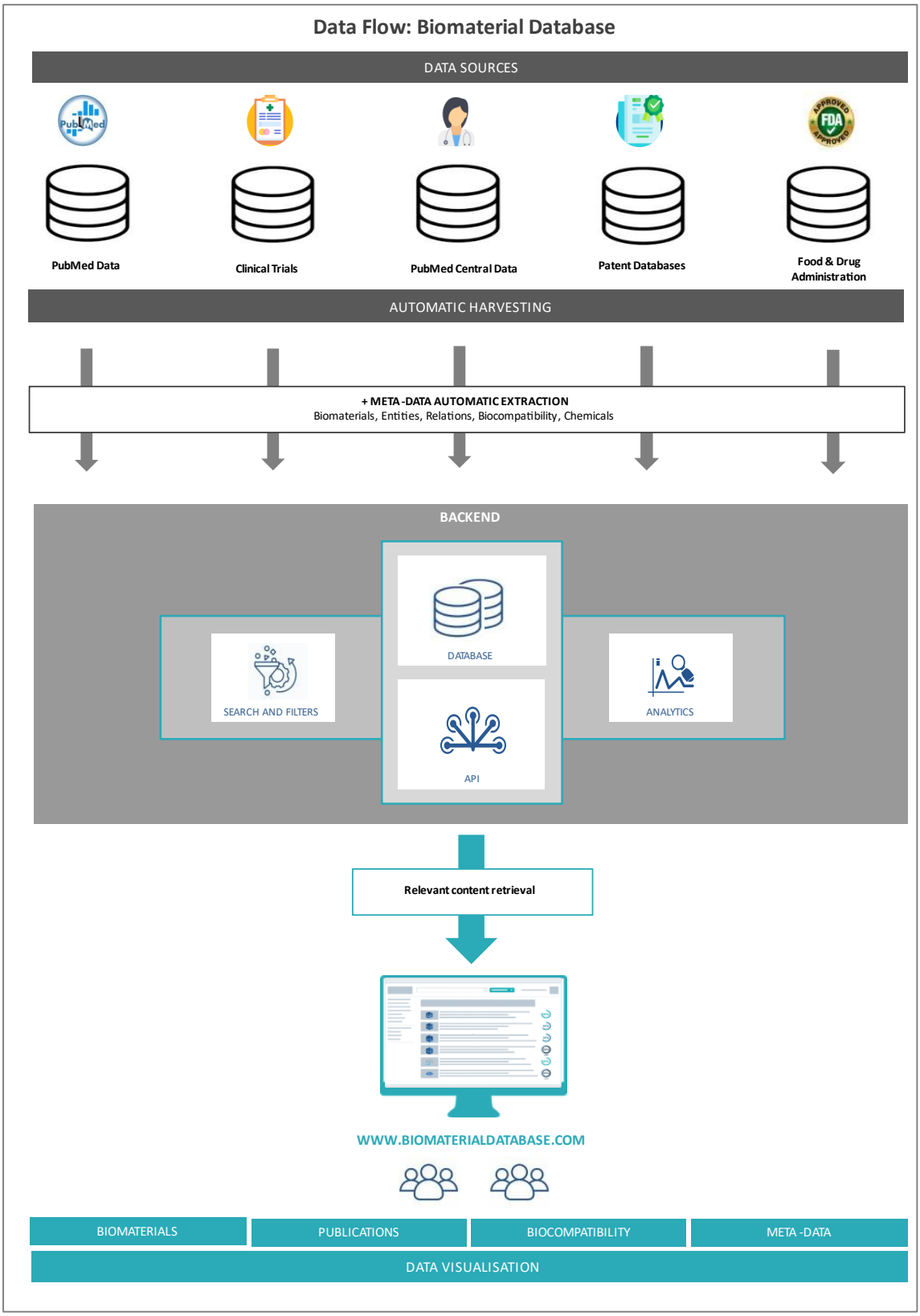



**Figure 3.** Dataflow from sources to frontend for the biomaterial database

Based on the survey results, the content sources that were of uttermost importance for the initial stages of the Biomaterial database included PubMed (relevant for 70% of respondents), Google (54%), Google Scholar (37%), Web of Science (29%), Literature (29%), Scopus (25%), University portals (13%), Science Direct (13%), Google Patents (8%), EuropePMC (4%), OVID (4%), Mendeley Datasets (4%), MHRA (4%), and FDA recently approved medical devices (4%).

In the case of the Biomaterial Database, its practical usage scenario has a clear research and development end-user focus. As a result, ensuring that the technical database design covers the interaction and processing of biomaterials, scientific literature content, together with the application of relevant content types (i.e., clinical trials, approved medical devices in the US market and their biomaterial compositions and patents), has an extreme relevance.

Moreover, several criteria were taken into account (or considered) for the content selection, such as legal aspects (licensing issues and access/redistribution restrictions), technical characteristics of automatic content retrieval, content formats, document types, and available metadata characteristics.

For the initial stages of the Biomaterial Database content, the main aim was to cover the content and process the biomaterials-relevant information from the PubMed database of biomedical scientific research abstracts and citation information. In addition, to cover full-text articles—a critical information source for more detailed biocompatibility study results of biomaterials—the PubMed Central database of full-text publications was considered.

### 2.2.2 Data Harvesting

Regarding task 2, the data harvesting and retrieval process is highly dependent on the underlying access options associated with each of the data sources that were processed and included into the Biomaterial Database. To provide a complete, understandable, and realistic scenario related to the harvesting strategies, concrete examples will be provided on the content retrieval and the data downloading techniques that were planned to be used for high priority content sources, using state-of-the-art web harvesting and web content pre-processing tools.

#### Data extraction strategies

In order to get the list of biomaterials-relevant domains, a crawler system was designed, capable of accessing and extracting the information contained in different websites. Applying a pre-processing step by removing the text marks and finally storing it. For the storage of the downloaded data, the guidelines of the International Internet Preservation Consortium (IIPC) were followed using the WARC (Web ARChive) file format, otherwise, if properly structured metadata is available (e.g Pubmed or PubmedCentral XML file structure), the alternative file types were also considered for data harvesting purposes. In addition, along with the extracted information from websites, a record of the crawler process/data harvesting was made, collecting for each domain processed, information such as the name of the domain, date of download, location, and name of the output files to facilitate subsequent processing, and information on possible access errors (non-existent domains, fallen pages). The record of this information will be stored in a database for easy and quick access. The architecture of this software system is modular, centralised on a database that stores the list of web addresses to be downloaded in order to subsequently provide them to a web crawler. Figure 4 shows the block diagram of the data harvesting system.

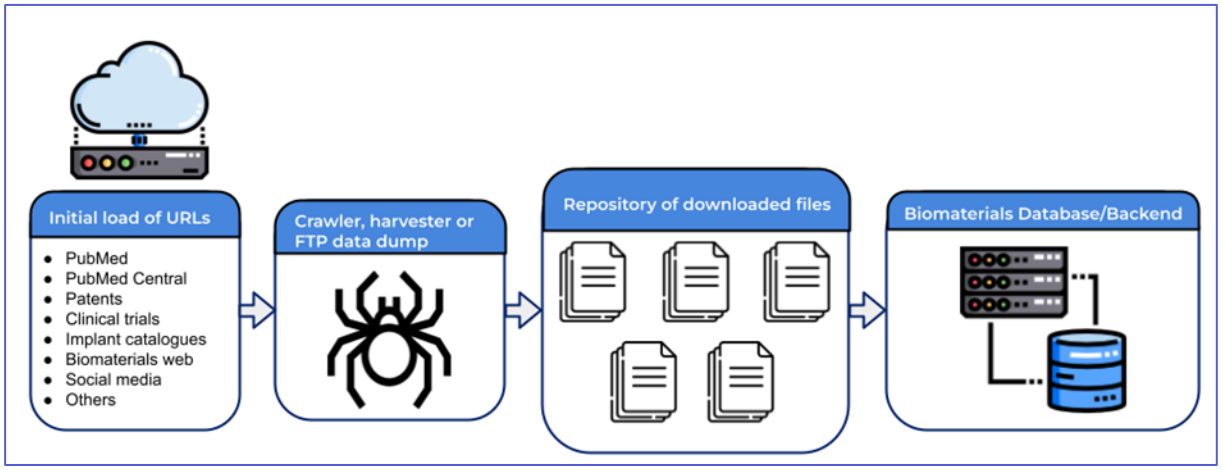



**Figure 4.** Data harvesting strategy for the Biomaterial database content

In order to access scientific literature content, including biomaterials research papers, a primary focus on free-access resources was planned. These resources were the content from PubMed, including citations and abstracts of research papers (currently in the order of 35 million records), and PubMed Central full text research articles, including biomaterials research papers or clinical case report publications where biomaterials are being used in clinical usage scenarios.

### Content Access

Two alternative content access strategies for the PubMed content were planned as follow: (1) a selective access of particular records using systematic access through web services offered by NCBI and (2) a database dump of the entire collection of PubMed through FTP access.

For the selective content access, the Entrez Programming Utilities (E-Utilities) resources were used, consisting of a public Application Programming Interface (API) that allows connectivity with all Entrez databases, including both PubMed and PubMed Central. This API will allow us to make complex, customised NCBI-supported queries to programmatically retrieve the resulting list of PubMed record hits.

In summary, the E-utilities that were planned to exploit, can be viewed as a set of server-side programs that allow to query PubMed and download search results programmatically, being the most common way to construct a Uniform Resource Locator (URL), a standard way to identify and access web pages, with the search query and use the command-line tool curl to download the results.

This selective content access strategy will also allow us to harvest citation outlinks to full text content, currently not included by the PubMed database bulk File Transfer Protocol (FTP) data access.

On the other hand, the second content access strategy is based on the design and implementation of other PubMed-based text mining solutions, such as DEBBIE (the first automatically curated database of biomaterial-related entities from PubMed abstracts), consisting of PubMed bulk download through its dedicated FTP servers. This allows users to download all of the PubMed data using yearly base releases together with periodic update archives for new publications included during the current publication year (i.e., from 2023 onwards). The FTP server offers data packages structured into different content folders, one of them covering the baseline release of PubMed (citation records in XML format until 2023). The second FTP folder covers new citations being periodically updated during the current release year of 2023.

Standard Python libraries were used for the extraction and handling of PubMed specific metadata and XML content. Mainly Beautiful Soup, a library that includes functionalities both for HTML and XML processing.

To specifically handle the processing, search, and exploitation of unstructured free text data elements contained in written scientific article titles and abstracts (as well as full text content in the case of PubMed central records), the text data elements were stored in a Elasticsearch, i.e., NoSQL Database environment for further data analysis.

### Initial pilot development

In order to design, implement and test the initial pilot Biomaterial Database, the first development stages were followed on a carefully selected set of biomaterials-relevant journals, based on the results and feedback from the initial survey of consortium biomaterials experts. This seed set of high priority journals was furthermore enhanced and extended directly through PubMed searches, journal and content metadata characteristics, including similar article features, Medical Subject Headings (MeSH) terms and keywords.

As a result, the first pilot implementation was focused on a 157 biomaterial-relevant publication journals set, hosted by PubMed, and partially also covered by PubMed Central. Moreover, to assure that this initial pilot implementation covers a substantial data space for building the Biomaterial Database, the resulting number of biomaterials relevant publication citations were quantified from this seed journal collection, resulting of 340.287 total records, ensuring that the initial database development scenario covers the main biomaterial classes. All of this initial exploratory content of relevance for the database design was extracted using Python in a dataframe structure, and saved in the biomaterials data repository.

To ensure that the Biomaterial Database also covers content related to the application of biomaterials being used in clinical scenarios (mainly as part of implants and devices), relevant information derived from clinical trial report was planned to be processed and included, focussing initially in the Clinicaltrials.gov (as well as the Spanish national clinical trials repository called REEC – Registro Español de Estudios Clínicos). For harvesting Clinicaltrials.gov content, consisting of a registry of clinical trials run by the United States National Library of Medicine (329,000 trials from 209 countries), a bulk-download option was used to retrieved the entire collection of XML formatted clinical trial documents offered by Clinicaltrials.gov. Then, data extraction and processing of XML content was handled by the Python package Beautiful Soup.

Finally, the third type of content that was planned to cover in the initial implementation of the Biomaterial Database corresponds to information of biomaterials in clinical use, that is being part of devices and implants. In this regard, currently the most relevant and comprehensive data source that can be accessed systematically is the U.S Food and Drug Administration Recently-Approved Devices website. This web resource provides information on medical devices that have been approved by the FDA, including their names, indications for use, and approval dates.

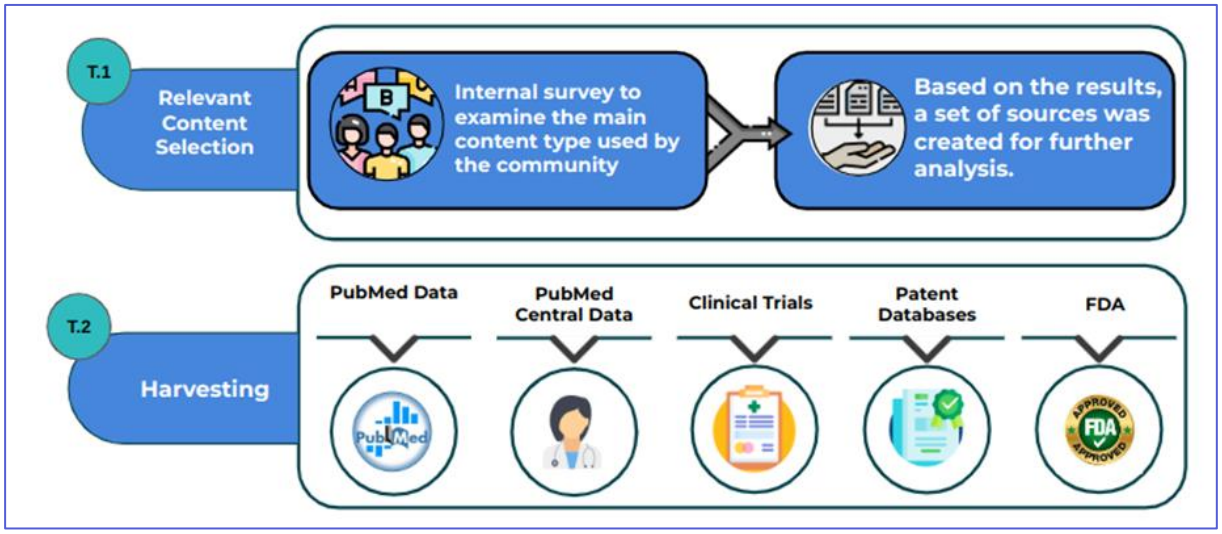



**Figure 5.** Data Sources and Harvesting

### 2.2.3 Data pre-processing and harmonisation

The next step concerning the pre-processing of the extracted data refers to Task 3 and considers that the process and integration of the heterogeneous data collections into the Biomaterial Database must be systematically integrated. As a result, this task does pose some technical challenges that need to be addressed.

The data pre-processing functionalities will involve cleaning the raw data, checking data consistency, as well as meta-data quality and completeness analysis, removing duplicates and standardising the data.

Data cleaning will involve the identification and correction of potential errors in the data, especially in terms of certain metadata elements (missing elements, element inconsistencies). We plan to deal with the duplicates by identifying them either through metadata element comparisons, or content-based record duplicate detection.

Moreover, standardising the data will also involve the transformation of raw data from heterogeneous formats into a standard document structure, suitable for automatic semantic processing by subsequent NLP components. The aim is to provide a higher level of abstraction of the data coming from the different sources, that is, a common schema for all sources (although there might be some specific properties from each one). During this data pre-processing step, which is referred to as data harmonisation, we will ensure the usage of a consistent metadata schema and format for data elements such as dates, names, source, etc.

## 2.3 Data entities, formats and types

As mentioned above, PubMed is a database of biomedical literature maintained by the U.S. National Library of Medicine (NLM) that contains over 32 million citations for biomedical literature, including articles, conference reports, books, and more. When downloading the data from PubMed, there are mainly two different formats available:

1. PubMed Abstract Format (PMID): This format provides a structured summary of the article's information, including the title, authors, abstract, publication details and MeSH terms. It is commonly used to retrieve a large number of article records from PubMed. In the present case, the data was downloaded as an XML format based on the National Library of Medicine's (NLM) Document Type Definition (DTD) for MEDLINE, which defines the structure of the XML file and includes the several fields of relevance:
    - Article title
    - Authors
    - Journal Name
    - Publication date
    - MeSH terms (Medical Subject Headings)
    - Abstract
    - PMID (PubMed ID)
    - DOI (Digital Object Identifier)
    - Publication type
    - Affiliations
    - Funding sources

2. PubMed Central (PMC) Full Text Format: This format includes the complete content of the article, including the abstract, full text, figures and tables, that was extracted in XML format. The PubMed Central is a free digital repository of full-text biomedical and life science journal articles. Not all articles in PubMed have this full-text available, but for those that satisfy this, the explained format provides access to the complete article content.

With respect to the MeSH terms provided by PubMed, these terms are a controlled vocabulary from PubMed used to index and categorise medical information into databases, covering a broad range of topics in the biomedical field, including anatomy, diseases, drugs, and medical procedures. It was developed and maintained by the National Library of Medicine, and is regularly updated to reflect changes and new developments in the field of medicine. From the initial implementation focused on a 157 biomaterial-relevant journals set hosted by PubMed and also partially covered by PubMed Central the most following MeSH term were extracted along with its respective occurrence:

**Table 3.** Most frequently extracted MeSH terms from PubMed

| MeSH ID | MeSH term | Freq. |
|---|---|---|
| D006801 | Humans | 155666 |
| D000818 | Animals | 84788 |
| D008297 | Male | 53307 |
| D005260 | Female | 48678 |
| D000328 | Adult | 27583 |
| D008875 | Middle Aged | 26991 |
| D008422 | Materials Testing | 26703 |
| D051379 | Mice | 25138 |
| D013499 | Surface Properties | 22286 |
| D001672 | Biocompatible Materials | 22282 |

## 2.4 Data examples and showcases

```
<!DOCTYPE ArticleSet PUBLIC "-//NLM//DTD PubMed 2.8//EN"
"https://dtd.nlm.nih.gov/ncbi/pubmed/in/PubMed.dtd">
<ArticleSet>
<Article>
<Journal>
<PublisherName>Public Library of Science</PublisherName>
<JournalTitle>PLoS One</JournalTitle>
<Issn>1932-6203</Issn>
<Volume>9</Volume>
<Issue>10</Issue>
<PubDate PubStatus="epublish">
<Year>2014</Year>
<Month>10</Month>
<Day>23</Day>
</PubDate>
</Journal>
<ArticleTitle>Correction: Statin Treatment in Hypercholesterolemic Men Does Not Attenuate
Angiotensin II-Induced Venoconstriction</ArticleTitle>
<FirstPage LZero="save">e112205</FirstPage>
<AuthorList>
<Author>
<CollectiveName>The PLOS ONE Staff</CollectiveName>
</Author>
</AuthorList>
<PublicationType>Published Erratum</PublicationType>
<ArticleIdList>
<ArticleId IdType="doi">10.1371/journal.pone.0112205</ArticleId>
<ArticleId IdType="pii">PONE-D-14-46356</ArticleId>
</ArticleIdList>
<Abstract>This corrects the article DOI: 10.1371/journal.pone.0103909.</Abstract>
<ObjectList>
<Object Type="Erratum">
<Param Name="type">pmid</Param>
<Param Name="id">25264877</Param>
</Object>
</ObjectList>
</Article>
</ArticleSet>
```

**Figure 6.** PubMed XML example file

# 3 Biomaterial Database: Processing methods, frameworks and visualisations

The Biomaterial Database will rely on advanced text data analytics and NLP solutions, tailored towards the necessities of the biomaterials research and development communities. At the document level topic modelling, text similarity, text classification, and clustering techniques will be applied. In the case of the previous resources such as DEBBIE [1], [2], [3], to triage literature for selecting publication records relevant for biomaterials, only basic supervised machine learning text classification approaches were used.

## 3.1 Processing methods

### 3.1.1 Topic modelling techniques

We plan to apply a topic modelling as NLP technique for extracting latent topics from the collection of biomaterials content hosted by the Biomaterial Database. It is a form of unsupervised learning, meaning that the algorithm is not given any prior information about the topics or categories that exist within the text. Instead, the algorithm identifies patterns of words and phrases that tend to occur together, and groups them into clusters or topics. In other words, this method can be used to explore a set of documents (corpus) by identifying different topics contained in the corpus.

There are several approaches to topic modelling, but two of the most widely used are Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). LDA is a probabilistic approach that assumes that each document is a mixture of different topics and each topic is a probability distribution over a fixed vocabulary of words. This method is by far the most popular approach for topic modelling. On the other hand, NMF is also a supervised learning technique which performs clustering as well as dimensionality reduction. It can be used in combination with a Term Frequency-Inverse Document Frequency (TF-IDF) scheme, a statistical measure that evaluates how relevant a word is to a document in a collection of documents that can be applied to perform this topic modelling.

In Python, there are several libraries available for topic modelling without requiring programming skills or knowledge of NLP techniques. These include Gensim, Scikit-learn, and NLTK. These libraries provide easy-to-use functions for pre-processing text, building topic models, and visualising the results.

### 3.1.2 Text similarity techniques

The NLP text similarity task involves measuring the similarity between two or more fragments of text, (words, sentences, or even entire documents). Meaning that its aim is to determine how close the content of each text is in terms of lexical or contextual meaning. Text similarity applications have a variety of purposes that includes document retrieval, plagiarism detection, and recommendation systems among others.

This technique is planned to be applied on the database content, based on different strategies that involve a morphological level (lexical similarity) and the contextual information of the text (semantic similarity). In terms of the lexical similarity, there are several algorithms that can be implemented in the development of the project, including Jaccard Similarity, Cosine Similarity and Euclidean Distance to mention the most common ones.

Each of these methods uses a different approach to quantify the similarity between documents, the method selection will depend on the specific case and the requirements of the application. For instance, the Jaccard Similarity calculates the similarity between two documents by comparing the number of common words they contain, while Cosine Similarity takes into account the frequency of words in each document and the orientation of their mathematical representation in space (text vectors). Euclidean Distance, on the other hand, measures the straight-line distance between two points in a multi-dimensional space, where each dimension corresponds to a word in the document.

The application of this technique can provide efficient and effective comparison of information contained within a large number of biomaterials research papers, identifying duplicate papers, finding similarities between papers in terms of their content, and clustering papers based on the field. These algorithms can also be used to compare the abstracts or full texts to match papers that are relevant for a particular research question, recommend papers to researchers based on their interests, or suggest new research directions based on the latest developments in the field.

The following pairs of sentences provide real examples of jaccard lexical similarity measures, computed from biomaterials article titles from the database:

Pair of sentences highly similar:

- **Sentence 1**: “The effect of the composition of hydrophilic base on pharmaceutical availability of prednisolone”.
- **Sentence 2**: “The effect of hydrophilic base on pharmaceutical availability of prednisolone complexed with beta-cyclodextrin.”
- **Similarity** (jaccard distance): 0.69

Pair of sentences with low similarity:

- **Sentence 1**: “Remarkably enhanced stability and function of core/shell nanoparticles composed of a lecithin core and a pluronic shell layer by photo-crosslinking the shell layer: in vitro and in vivo study.”
- **Sentence 2**: “In vitro study of hemodynamic treatment improvement: Hunterian ligation of a fenestrated basilar artery aneurysm after coiling.”
- **Similarity** (jaccard distance): 0.15

### 3.1.3 Text classification and clustering

Text classification with clustering involves using methods to categorise text data into predefined classes or clusters. The goal of text classification is to automatically assign each document in a collection to the correct class or cluster based on its content and context. In other words, it allows for the discovery of natural groups or clusters in the data without any prior knowledge about the number or structure of the classes.

In the context of this project, text classification and clustering algorithms were proposed to be applied for the categorization of research papers based on content, meaning, and context. This can allow the discovery of natural groups or clusters of papers, which can provide valuable insights into the structure of the biomaterials research community, and the relationships between researchers, institutions, and biomaterials.

In relation to the text classification, the first step will consist of creating a set of labelled data according to predefined labels. The annotation will be done in free text abstracts and the MeSH terms extracted from PubMed using the Prodigy software.

In the case of BiomatDB, Prodigy will display to the experts some biomaterials samples to be classified based on specific labels or criteria, along with a set of instructions or guidelines to annotate them properly. Once the annotation results be collected and organised, the resulting labelled data can be implemented in the text classifier development. By using prodigy for annotation, time and effort will be saved compared to manual annotation methods, in addition to the active learning component that may be helpful to identify the most important samples to annotate. Figure 7 shows an illustrative example of the different predefined labels to the annotator in Prodigy.


**Figure 7.** Example of the annotation display in prodigy concerning the biomaterial study

Four different annotation cases have been proposed, three of them are directly related to the information provided in the abstract itself under the following classification: Biomaterial type, study type and biocompatibility & features. The fourth case is related to the classification of the MeSH terms. The labels for each case are presented as follow:

**Table 4.** Predefined prodigy annotation labels

| Biomaterial type | Study type | Biocompatibility & features | MeSH terms |
|---|---|---|---|
| ▪ Not<br>▪ Other<br>▪ Metallic<br>▪ Polymeric<br>▪ Ceramics<br>▪ Natural<br>▪ Bioink<br>▪ Glass<br>▪ Composite<br>▪ Blend<br>▪ Alloy<br>▪ Synthetic<br>▪ Unclear | ▪ None<br>▪ Unclear<br>▪ Clinical study<br>▪ In vivo<br>▪ Ex vivo<br>▪ In vitro<br>▪ Pre clinical<br>▪ Omics<br>▪ Other | ▪ No biocompatibility evidence<br>▪ Not biocompatible<br>▪ Biocompatible<br>▪ Biological compatibility<br>▪ Functional check<br>▪ Mechanical behaviour<br>▪ Physical shape<br>▪ Physical features<br>▪ Manufacturing techniques<br>▪ Biodegradation | ▪ Biomaterial<br>▪ Medical Application<br>▪ Study Type<br>▪ Manufactured Object<br>▪ Associated Biological Process<br>▪ Tissue Type<br>▪ Structure<br>▪ Effect on Biological System<br>▪ Cell Type / Line<br>▪ Material Processing<br>▪ Adverse Effects<br>▪ Animal Model<br>▪ Manufactured Object Features<br>▪ Biologically Active Substance<br>▪ Research Technique<br>▪ Biocompatibility<br>▪ Other material/substance<br>▪ Other |

### 3.1.4 Named Entity Recognition (NER)

Named entity recognition (NER) systems based on Transformers and biomedical language models will be used to automatically detect mentions of biomaterial-relevant concepts or classes from the heterogeneous document types, stored in the Biomaterial Database. Moreover, detected concept mentions will be automatically mapped to DEBBIE biomaterials ontology concepts, in addition to relevant structured vocabularies, such as UMLS, MeSH, and SNOMED CT (this normalisation is discussed further in the next subsection).

The NER process, along with the posterior linking of terms to an ontology (usually referred to as NER+L), and the relation extraction (hyponyms) are the main components of Task 5 (T.5). The resulting automatic semantic annotations of the Biomaterial Database content will serve to enable more robust search queries and strategies for biomaterials information.

We will use several NER systems to recognise all the proposed entities. Some of them will be simple dictionary lookups, for instance to the Debbie and SnomedCT terminologies, while others will be new or existing neural network-based models (mostly available in scispaCy and Hugging Face). Note that we can have some redundancy to provide more robustness to predictions, that is, multiple models looking for the same kind of entity.

For the new models that will be implemented, pre-trained transformer-like models will serve as base models for our specific NER systems. These models will be fine-tuned with the annotations of the corresponding entities to automatically detect biomaterials-relevant concept types, including entity classes for material, chemical composition, anatomical entities, cell types, cell lines, study type, and certain biomaterial attributes.

Concretely, the biomaterials concept NER predictions will be initially modelled as a token classification problem. As a result, a text span for each of the detected biomaterial-related concept mentions, along with its entity type (e.g., cell, protein, material, etc.), is obtained.

Like in most of the state-of-the-art approaches, we will attach a model head (normally a dense layer with some regularization layer after) on top of a public pre-trained biomedical language model, such as PubmedBERT [4], ClinBERT [5], BioBERT [6] or some of the pre-trained models built by, or in conjunction with, the BSC.

The parameters or weights of this model head are randomly initialised and then adjusted on a downstream task, which in this case will be to recognise the different named entities. Many iterative experimentations will take place in this step, regarding the various base models, configurations and hyperparameters of the neural network, to ensure the best performance of the model.

An advantage of this paradigm (pre-trained + fine-tuning), which is extensively used in the literature, is that it allows us to leverage the knowledge in the base model. Additionally, it enables us to need far less manually annotated data and computation resources for training, when compared to training a model from scratch, thus achieving much better results.

In order to run large-scale Named Entity Recognition systems, certain text pre-processing and post-processing steps will be applied before and after the NER models, respectively. The input will be processed considering the context of the mentions, i.e., close words and cross-sentence information available in consecutive sentences of the input documents. For the scispaCy-based models, the pre- and post-processing of the large-scale data will be based on components offered by the spaCy library, including sentence splitting and word tokenization.

At the technical level, we will employ primarily spaCy and Python Deep Learning libraries (PyTorch and HuggingFace) to develop the automatic entity extraction models for biomaterials. Once the models present high enough performance (which is described in the next section, 3.2), we plan to scale these two components to process the entire biomaterials-related PubMed, PubMed Central, and other content sources included within the Biomaterial Database.

Finally, to address current limitations in terms of biomaterials attributes not covered well either by existing named entity resources or controlled vocabularies/ontologies, we will exploit automatic keyphrase extraction technologies with BERT-based [18] Transformers and Noun Phrases, by means of multilingual KeyBERT [8] or PatternRank [9]. The application of these technologies will enable automatic ontology concept candidate detection and semantic indexing of biomaterials content.

### Initial NER extraction

As initial validation of the proposed NER solution, we have run all NER models on a subset of 1000 abstracts from PubMed. This provides us the first indication on the distribution of data we are dealing with, that is, the frequency of recognized entities.

From this initial NER extraction, we will have to select and keep just the relevant entities. The entities found by the models will be used as suggestions or pre-annotations for annotating the data. More information of each model/component can be found in section 3.3.

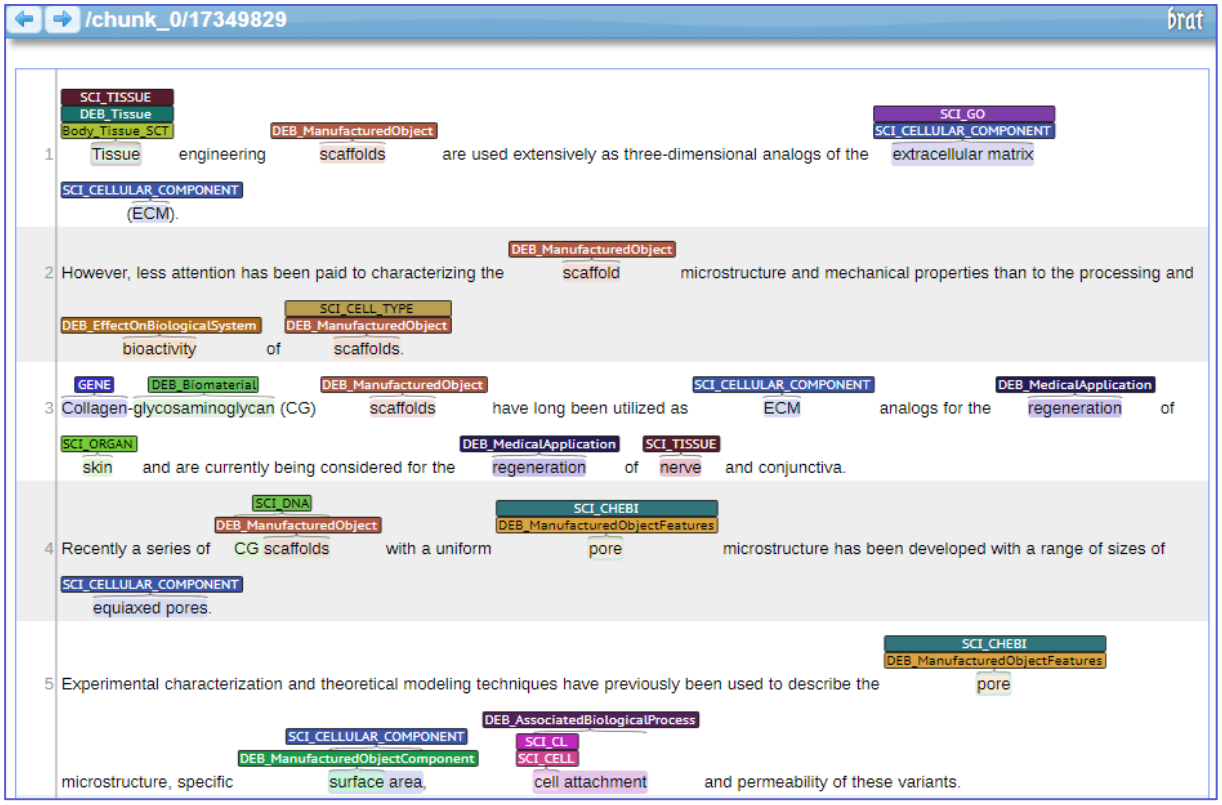



**Figure 8.** Brat annotation tool with the initial pre-annotations from the models

**Table 5**. Frequency of named entities in the initial subset of 1000 PubMed abstracts

| Model | Label | Number of documents containing entities | Number of entities found |
|---|---|---|---|
| **SnomedCT look-up** | | | |
| **SnomedCT dictionary look-up** | Biomedica Device | 110 | 124 |
| | Body Tissue | 256 | 336 |
| | Cell Structure | 276 | 392 |
| | Implant | 35 | 39 |
| | Material | 381 | 472 |
| | | | |
| **DrugProt** | | | |
| **DrugProt Chemical** | Chemical | 641 | 4937 |
| **DrugProt Gene** | Gene | 419 | 3137 |
| | | | |
| **DEBBIE look-up** | | | |
| **DEBBIE dictionary look-up** | Adverse Effects | 180 | 380 |
| | Architectural Organization | 45 | 102 |
| | Associated Biological Process | 262 | 674 |
| | Biologically Active Substance | 119 | 292 |
| | Biomaterial | 435 | 1635 |

| | | | |
|---|---|---|---|
| | Biomaterial Type | 210 | 562 |
| | Cell | 254 | 526 |
| | Degradation Features | 97 | 155 |
| | Effect On Biological System | 232 | 515 |
| | Manufactured Object | 360 | 1272 |
| | Manufactured Object Component | 234 | 481 |
| | Manufactured Object Features | 390 | 978 |
| | Material Processing | 473 | 1412 |
| | Medical Application | 280 | 604 |
| | Research Technique | 450 | 1328 |
| | Shape | 52 | 74 |
| | Species | 26 | 30 |
| | Structure | 338 | 908 |
| | Study Type | 265 | 461 |
| | Tissue | 541 | 1591 |
| | | | |
| **scispaCy** | | | |
| **en_ner_craft_md** | ChEBI[1] | 681 | 3183 |
| | CL[2] | 518 | 2753 |
| | GGP[3] | 409 | 1536 |
| | GO[4] | 258 | 474 |
| | SO[5] | 391 | 867 |
| | Taxon | 463 | 1130 |
| **en_ner_jnlpba_md** | Cell Line | 235 | 461 |
| | Cell Type | 505 | 1964 |
| | DNA | 377 | 793 |
| | Protein | 670 | 3230 |
| | RNA | 17 | 32 |
| **en_ner_bc5cdr_md** | Disease | 659 | 2879 |
| | Chemical | 677 | 3596 |
| **en_ner_bionlp13cg_md** | Amino Acid | 61 | 75 |
| | Anatomical System | 6 | 9 |
| | Cancer | 324 | 762 |
| | Cell | 682 | 3929 |
| | Cellular Component | 413 | 1026 |
| | Developing Anatomical Structure | 2 | 4 |
| | Gene Or Gene Product | 626 | 2882 |
| | Immaterial Anatomical Entity | 93 | 120 |

[1] Chemical Entities of Biological Interest
[2] Cell Ontology
[3] Gene or Gene Product
[4] Gene Ontology
[5] Sequence Ontology

| | | | |
|---|---|---|---|
| | Multi Tissue Structure | 431 | 1244 |
| | Organ | 316 | 862 |
| | Organism | 571 | 1447 |
| | Organism Subdivision | 155 | 219 |
| | Organism Substance | 206 | 471 |
| | Pathological Formation | 248 | 453 |
| | Simple Chemical | 585 | 2628 |
| | Tissue | 623 | 2255 |

### 3.1.5 Automatic Acquisition of Hyponyms from biomaterials abstracts

The idea behind this subtask is to automatically identify a set of lexico-syntactic patterns that are easily recognizable and can occur frequently across the different biomaterial's abstracts. In the context of the project, this will consist of hyponyms pairs detection based on predefined patterns known as "Hearst Patterns". The implementation of this detection over the text data will help to enrich the ontology, generating biomaterial candidates during the annotation process, mapping the extracted relation within the brat annotation or even generating knowledge graphs according to the relations extracted.

The following relations (<connector>, <hyperonym>, <hyponym>), examples were extracted from the pilot implementation based on more than 300k PubMed works study.

- *'like', biomaterials, collagen*
- *'such_as', metals, titanium*
- *'such_as', biomaterials, hydrogels*
- *'for_example', tissue engineering applications, biomaterials*
- *'especially', calcium phosphates, tricalcium phosphate*
- *'such_as', agents, antracyclines*
- *'such_as', agents, cyclofosfamide*
- *'such_as', therapies, imatinib*
- *'include', keratoplasty, modification*
- *'particularly', nanomaterials, metal oxides nanoparticles*
- *'such_as', applications, tissue engineering*

### 3.1.6 Metadata and entity normalisation

Once the source data has been cleaned and standardised, and the relevant metadata has been automatically extracted by means of NLP techniques, a normalisation of terms will take place. In the case of the Biomaterial Database, this process will require mapping relevant data elements (both in terms of actual metadata content as well as NLP-extracted automatic semantic annotations) to corresponding biomaterials-controlled vocabulary concepts. For example, if the data element refers to information about diseases, the diseases should be mapped to a structured vocabulary, such as UMLS, the International Classification of Diseases (ICD), SNOMED-CT, or DEB Ontology.

The metadata and entity normalisation makes it easier to compare and analyse data from different sources, empower query expansion during database searches, and ensures semantic interoperability across multiple content sources sorted in the Biomaterial Database.

This will enable the system to recognize that multiple mentions (expressed differently) are actually referring to the same concept, and exploit the knowledge or hierarchies derived from these ontologies.

The first strategy that will be implemented will explore vectorizing the biomaterials concept mentions, as well as the targeting ontology terms (or synonyms, descriptions…), and then finding the best candidate match by computing the vector similarity between the biomaterial-related mentions and the ontology terms. Moreover, we plan to use an ensemble of term frequency–inverse document frequency (TF-IDF), character-n-gram-based approaches, and multilingual embeddings using SapBERT to offer more advanced biomaterials concept linking to target ontologies.

## 3.2 Iterative improvement of guidelines, annotations and machine learning models

This section refers to Task 6 (T.6), which describes the methodology to be implemented to iteratively refine the resources generated in Tasks T.4 and T.5, namely the corpus (guidelines and annotations) and the trained models on that corpus.

### 3.2.1 Creating the corpus

Developing quality NLP systems adapted to the biomaterials domain requires annotating useful data for both training and evaluation. In the case of textual data, annotating means associating the text with meta-data that describes some aspects of it which can be transformed into machine-readable input and used to create artificial intelligence (AI) systems. For each of the NLP tasks described above, the annotation is different.

For instance, for a text classification task, one or more labels are assigned to full documents. In contrast, for Named Entity Recognition, the annotation consists of matching specific text spans—rather than entire documents—with semantic classes (such as chemical object, disease, person or organisation).

Every annotation process has different steps that should be carefully considered and planned before starting the project, in order to minimise costs and efforts. Roughly, these include: task definition, data selection, text annotation, and dataset postprocessing.

The first part of an annotation project involves defining the annotation task. This includes deciding what is going to be annotated, in which type of texts, and how many documents, by whom, besides other basic aspects, like the most adequate annotation tool and a general literature review. Most of these decisions will be documented in a document called “annotation guidelines”. These guidelines are a keystone document that gathers general information about the project, together with the criteria to be used for the annotation.

Next, the documents to be annotated need to be selected. It is important to choose texts that accurately reflect the linguistic and semantic variety of the type of documents that the final systems will be used on. Document selection can be done both manually and semi-automatically, by using intelligent selection methods such as text similarity.

The text annotation step has two well-defined parts: guideline refinement and annotation escalation. The former consists of iteratively improving the guidelines document so that the criteria are well-defined, unambiguous, and adjusted to the project aim.

One way to do this is through parallel annotation, a process in which several small groups of documents are annotated between two or more people.

Parallel annotated documents are compared using consistency metrics (such as inter-annotator agreement or IAA) to compare similarity between annotators. Based on the differences between the annotators, the guidelines are refined with both reworked and new rules and examples. When a high score is reached, it is considered that the guidelines clearly describe the annotation criteria, and the rest of the document collection can be annotated. The expected IAA score depends entirely on the difficulty of the task. Robust annotation guidelines are important to ensure consistency and coherence in the resulting dataset.

Once a sufficient IAA has been reached, the annotators can start working on the rest of the selected documents—annotation escalation. During this part, the role of the annotation managers is to solve possible questions, manage annotator workload, and continue refining the guidelines with new examples and decisions, in addition to help integrate intermediate models trained on the already annotated data. These intermediate systems can be used to enrich new texts with suggestions such that annotators don't have to work from scratch, accelerating the different annotation process. In other words, analysing the outputs of these models on already labelled documents can also inform the annotation process in a feedback loop, by locating errors in the annotated data and/or underrepresented data types.

Each partner will provide, at least, one annotator with expertise of biomaterials who will help develop new guidelines adapted to the domain and manually label new pre-annotated texts with some of our existing systems' predictions. Depending on the specific task or type of annotation, we will employ a different annotation tool. For instance, for text classification we will utilise the Prodigy annotation tool while Brat will be our preferred option for NER.

### 3.2.2 Validating the model

The final step is the evaluation of the models. In this part, the predictions of the models are assessed and checked through automatic and manual revisions. By using a subset of the manually annotated samples (not seen by the model during training), we can validate the model's performance. On the other hand, an exploratory analysis in the predictions (both per-case and general statistics), will be conducted to better understand the cases where the model performs worse (alongside its strengths). In case of non-optimal results, the annotations and its guidelines will be revised and/or more annotations will be included in the dataset, apart from experimenting with the model's hyperparameters and architecture. Note that it is always possible to go back and make another iteration to improve the models.

Overall, although the annotation process might be seen as time-consuming and expensive, manually annotated data offers several benefits in terms of accuracy, performance, and even traceability that make the investment worthwhile.

## 3.3 Annotation tools, software, and models

As already stated, several models will be used to automatically extract biomaterial-useful information. The nature and complexity of these models will range from simpler regex or dictionary look-ups to complex transformer-based neural networks. For the new machine learning-based models that will be implemented, we will use annotation platforms to produce the labels needed, namely Prodigy and Brat.

On the other hand, to train and integrate these models within the system pipeline (pre- and post-processing, inference once trained, etc.), different libraries and frameworks will be used, which will be detailed below. Note that we will use the Python implementation for all libraries, and that the solution will follow a modular design, so that they can all be easily integrated.

### 3.3.1 spaCy

spaCy [10] is is an open-source library for natural language processing, developed by Explosion AI. It provides efficient and scalable tools for different NLP tasks, such as tokenization, part-of-speech tagging, named entity recognition, and dependency parsing. spaCy focuses on performance, offering fast and accurate processing capabilities. It is widely used for building production-ready NLP applications and is known for its ease of use, extensibility, and integration capabilities. It will be used to create an NLP pipeline and integrate the look-up dictionaries (SnomedCT, DEBBIE, etc.) and scispaCy models within this pipeline.

### 3.3.2 scispaCy

scispaCy [11] is an extension of the spaCy library, specifically designed for scientific and biomedical text processing. It offers pre-trained models and tools tailored to the specific needs of the biomedical domain. scispaCy provides capabilities for entity recognition, relationship extraction, abbreviation expansion, and other tasks commonly encountered in biomedical literature analysis. It is developed by the Allen Institute for AI and is widely used in biomedical research and healthcare-related applications.

The scispaCy models are trained on data from a variety of sources. In particular, the following were used:

- The GENIA 1.0 Treebank, converted to basic Universal Dependencies using the Stanford Dependency Converter.
- word2vec word vectors trained on the Pubmed Central Open Access Subset.
- The MedMentions Entity Linking dataset, used for training a mention detector.
- Ontonotes 5.0 to make the parser and tagger more robust to non-biomedical text. This dataset is not publicly available.

We will use some of the already trained models in scispaCy to detect entities of special interest and link them to the corresponding ontologies.

### 3.3.3 PyTorch

PyTorch [12] is an open-source deep learning framework developed by Facebook. It allows researchers and developers to build and train neural networks easily. PyTorch is known for its simplicity and flexibility, making it one of the preferred choices for many in the research community.

It is the default underlying deep learning framework for Hugging Face, which already provides a higher-level abstraction for training and evaluating neural networks. Therefore, specific use of this library will not be necessary, except if low-level experimentation and customization of the neural networks are needed.

### 3.3.4 Hugging Face

Hugging Face [13] is an organisation focused on deep learning with special emphasis on NLP. They provide a range of open-source libraries, tools, and pre-trained models that facilitate various NLP tasks. Hugging Face's most popular library, *Transformers*, offers state-of-the-art pre-trained models and tools for tasks like text classification, named entity recognition, and machine translation. They also maintain a large repository of pre-trained models that can be easily fine-tuned and integrated into NLP pipelines.

Hugging Face will be the primary source of pre-trained models for fine-tuning in case we need to train any new model. On the other hand, the pre-trained models (not fine-tuned) can be integrated to the spaCy pipeline thanks to the spacy-transformers library, where they can be fine-tuned.

### 3.3.5 Tensorflow (Keras)

TensorFlow [14] is a popular machine learning framework developed by Google. It provides a set of tools and resources for building and deploying machine learning models. With TensorFlow, developers can create and train neural networks for tasks like image recognition, language processing, and more. One of the notable features of TensorFlow is its compatibility with Keras, a high-level neural networks API.

The DrugProt models, trained on the homonymous corpus defined in the Biocreative VII Track 1[6] are capable of detecting mentions of chemicals and genes. They were built using Tensorflow.

### 3.3.6 Brat

Brat [15] is an an open-source web-based annotation tool used for manual annotation of text data. Brat provides a user-friendly interface for annotating entities and their relations in text documents. It allows researchers and annotators to collaboratively label data and create annotated corpora for training machine learning models. Brat is highly customizable and supports various annotation schemas and workflows.

Brat will serve as a visualisation and annotation tool for developers and expert annotators, to show and modify the named entities and the relations between them, if any. Here, we can display the different sources of annotations (i.e., the ground truth and the predicted entities of the models), thus, providing the users with useful data to check, validate, or modify the annotations. An example of its interface is shown in Figure 8.

#### Prodigy

Prodigy [18] is an annotation tool developed by the creators of the spaCy natural language processing (NLP) library that allows the user to annotate and label various types of data. It is designed to be used in the context of machine learning, specifically for creating and improving training data sets for machine learning models. Prodigy provides an intuitive interface for data annotation, and it also

[6] https://zenodo.org/record/5119892

includes an active learning component that helps users to identify the most important data to annotate. Although Prodigy can be easily integrated with spaCy to train a model in an active learning manner, it is not a full-fledged machine learning framework per se, but rather a specialised tool for data annotation and labelling.

## 3.4 Database management system initial design

When developing the Biomaterial Database several key pre-requirements have to be characterised involving the following steps:

- **Data model definition**: This step begins by defining the entities along with their attributes to be stored in the database. For the biomaterials database, these will include information like the name, description, composition, properties, and source of the biomaterials. In our case, the proposed model is a "document model" which stores the data as a document containing the different attributes and fields.
- **Identify the relationships**: It is important to determine the relationships between the entities, one-to-many, many-to-many, or one-to-one relationships. For example, a biomaterial may have multiple properties and each property can be associated with multiple other biomaterials. Relation types can be based on document-level associations derived from the metadata information or co-occurrences within the document itself (e.g., biomaterial concept co-mentions, as is the case of the DEBBIE resource or qualified relation types like biomaterial-has-property, biomaterial-has-chemical-composition, etc.).
- **Choose a database management system (DBMS)**: In this case, DBMS is selected as the best suitable option regarding the project requirements. Common options include SQL-based systems, such as MySQL or PostgreSQL, or NoSQL systems such as MongoDB, Cassandra, or Elasticsearch. In this case, Elasticsearch will be used because of its ability to handle large volumes of data and complex querying. Table 6 specifies the main reasons of using Elasticsearch
- **Import the data**: Once the data model is defined and the DBMS is chosen, the next step is to import the data into the database. Depending on the size and complexity of the data, this can be done manually or with the help of an automated tool.
- **Index the data**: Indexing the data is crucial to ensure fast search and retrieval of information, and involves creating indexes on the most commonly queried fields in the database.
- **Develop the framework**: Finally, develop the framework for the biomaterials database, including the user interface and query system. This framework should be intuitive and easy to use, with a user-friendly interface for data entry and search.

**Table 6.** Main advantages of using Elasticsearch as DBMS

| Reason | Explanation |
|---|---|
| Full-text search capabilities | Elasticsearch is designed for full-text search, making it easy to search and filter through large volumes of data. In a biomaterials database, this would allow users to quickly and easily find specific biomaterials based on their properties, composition, or other characteristics. |
| Scalability | Elasticsearch is a distributed system that can scale horizontally across multiple nodes, making it easy to add more resources as needed. This would allow a biomaterials database to handle large volumes of data and support multiple users concurrently. |
| Speed | Elasticsearch is optimised for performance and can return search results in real-time, making it ideal for real-time analytics and data visualisation. This would allow users of a biomaterials database to interact with the data and generate insights quickly and efficiently. |
| Rich query capabilities | Elasticsearch supports a variety of query types, including Boolean, fuzzy, and proximity queries, as well as filtering and aggregations. This allows for complex queries and data filtering, making it easier to find the information needed in a biomaterials database. |
| Flexible data modelling | Elasticsearch provides flexible data modelling options, including support for nested and multi-valued fields. This would allow for a wide range of data types and structures to be stored in a biomaterials database. |

## 3.5 Interactive visualisation frontends and techniques

To interact with the Biomaterial Database, an interactive user interface compromising (semantic) search functionalities, visual analytic components and statistics about the data will be provided. The implementation step will make use of state-of-the-art frameworks and techniques to enable the generation of meaningful insight. To make an informed decision on which frameworks and techniques will be used, an analysis of existing libraries and approaches has been carried out (see Figure 9).

During the analysis, various solutions have been gathered and compared in a two-step process. First, the “must have requirements” were used as knock-out criteria (K.O. Criteria) to determine which solutions shall be analysed further. For this step, a decision matrix was used. The solutions were analysed based on the solutions’ documentation, web information and forums.

In a second step, the “should-have requirements” were used as criteria. These criteria were weighted according to importance. Thus, a priority ranking was established. For example, the requirement “the system should be able to read data in JSON format” was ranked as more important than the requirement “the system should offer English and German documentation”.

| Name | URL | Open Source | Canvas | SVG | Built-in Charts | 8+ Chart Types | Extendable to Custom Charts | Supports Modern & Older Browsers | Extensive Documentation | Notes |
|---|---|---|---|---|---|---|---|---|---|---|
| ChartJS | https://www.chartjs.org | x | x | x | x | x | x | x | x | There are 8 different types of charts out of the box (Line, Bar, Radar, Doughnut and Pie, Polar Area, Bubble, Scatter, Area), plus t |
| Google Charts | https://developers.google.com/chart | x | x | x | x | x | x | x | x | Charts are rendered using HTML5/SVG to provide cross-browser compatibility and cross-platform portability to iPhones, iPads, a |
| FusionCharts | https://www.fusioncharts.com/ | | | x | x | x | x | | x | Very comprehensive chart library. Has more than 100+ charts and 1400+ maps. Charts are rendered using HTML5/SVG and VML |
| pChart | http://www.pchart.net/ | x | | | x | x | | | x | |
| ChartLogix | https://github.com/elikhalouf/Chartbypho/blob/ | x | | | x | | | x | x | |
| D3 | https://d3js.org/ | x | x | x | x | x | x | x | x | It has unique features like dynamic properties, Enter and Exit, powerful transitions, and syntax familiarity with jQuery. Charts in |
| Chartist.js | https://gionkunz.github.io/chartist-js/ | x | | x | | | x | | x | It uses SVG to render the charts. It can be controlled and customized through CSS3 media queries and Sass. It has very modern a |
| Recharts | http://recharts.org/ | x | | x | | x | | x | x | Recharts is a composable charting library for building charts with decoupled, reusable React components. It's built on top of SVG |
| n3-charts | http://n3-charts.github.io/line-chart/ | x | x | | | x | x | | x | n3-charts is built on top of D3.js and Angular. |
| Ember Charts | https://opensource.addepar.com/ember-charts/ | x | | x | x | | | x | x | It provides time series, bar, pie, and scatter charts that are easily customizable. It uses SVG to render charts. |
| ZingChart | https://www.zingchart.com/ | | x | x | | x | x | x | x | ZingChart offers more than 35 responsive chart types and modules, which can also be showing data in real-time. They can be sty |
| Highcharts | https://www.highcharts.com/ | | | x | x | x | x | x | x | There are wrappers for most popular languages (.NET, PHP, Python, R, and Java) and frameworks (Angular, Vue and React), and f |
| Flot | https://www.flotcharts.org/ | x | | | | | x | | | Flot supports lines, points, filled areas, bars and any combinations of these. It's also compatible with older browsers — way back |
| amCharts | https://www.amcharts.com/ | x | x | x | x | x | x | x | x | amCharts uses SVG to render charts that work in all modern browsers. It provides integrations with TypeScript, Angular, React, V |
| EJSCharts | http://www.ejschart.com/ | x | x | | | | | | | |
| uvCharts | https://imaginea.github.io/uvCharts/ | x | | x | x | | x | | | uvCharts are rendered using SVG and HTML5, so it works only with modern browsers. |
| Plotly.js | https://plot.ly/javascript/ | x | x | x | x | x | x | x | x | Plotly.js supports 20 chart types, including SVG maps, 3D charts, and statistical graphs. It's built on top of D3.js and stack.gl. |
| CanvasJS | https://canvasjs.com/ | | | x | x | x | | | x | CanvasJS can be integrated with popular frameworks (Angular, React, and jQuery) and server-side technologies (PHP, Ruby, Pyth |
| TOAST UI Chart | https://ui.toast.com/ | x | x | | | | x | | x | It includes all common chart types and maps, which can be customized with user-defined themes. The charts can be also combi |
| AnyChart | https://www.anychart.com/ | x | x | x | x | x | x | x | x | AnyChart is a lightweight and robust JavaScript charting library with charts designed to be embedded and integrated. AnyChart |
| JSCharting | https://jscharting.com/ | | | | | x | | x | | JSCharting chart library supports a large number of chart types including maps, gantt, stock and others that often require separa |
| KOOLCHART | https://www.koolchart.com/ | | | x | x | | | x | x | KoolChart is an HTML 5 canvas-based JavaScript charting library. A mapping and grid product is also available. |
| Dygraphs | http://dygraphs.com/ | x | x | | | | x | | x | Dygraphs is an open-source JavaScript charting library best suited for extremely large data sets. It is interactive out of the box, w |
| NVD3 | http://nvd3.org/ | x | | x | x | | | | | It lets you handle complex data sets and create advanced visualisations. |
| jQuery Sparkline | https://omnipotent.net/jquery.sparkline/ | x | | | | | | x | x | |
| Sigma.js | http://sigmajs.org/ | x | | | | | x | | x | Sigma's libraries and plugins pack in an enormous amount of interactive settings. |
| Morris.js | https://morrisjs.github.io/morris.js/ | x | | | | | x | | x | Morris is a lightweight library based on jQuery and Raphael, that provides simple, clean line, area charts, bar, and donut charts. |
| Cytoscape.js | http://js.cytoscape.org/ | x | | | x | | x | | x | Cytoscape.js is an open source, fully featured graphing library, written purely in JavaScript. |
| C3.js | https://c3js.org/ | x | | | x | | | x | | C3.js provides another way around the D3 learning curve by wrapping the code required to construct the entire chart. C3 lets yo |
| Rickshaw | https://github.com/shutterstock/rickshaw | x | | x | x | | | | x | |
| Cubism.js | https://square.github.io/cubism/ | x | x | | x | | | | x | It renders incrementally, using Canvas to shift charts one pixel to the left at a time. Cubism's horizon charts make better use of v |
| Plottable.js | http://plottablejs.org/ | x | | x | | | | | x | |
| Charts4PHP | https://www.chartphp.com/ | | x | x | x | | | x | x | Charts 4 PHP Frameworkit is possible to create interactive database driven HTML5 dashboard using mySQL or any other PDO su |
| GoJS | https://gojs.net/latest/index.html | x | | x | | | x | x | x | GoJS is a JavaScript diagramming library for interactive flowcharts, org charts, design tools, planning tools, visual languages. |
| dimple | http://dimplejs.org/ | x | | x | x | | x | | | dimple is an object-oriented API for business analytics powered by D3. |

**Figure 9.** Decision Matrix with decision criteria

Based on this analysis, the libraries which are considered are D3.js, Chart.js for the visualisations and Bootstrap as CSS-library.

### 3.5.1 Data-Driven Documents (D3)

D3.js [16] is a widely used open-source JavaScript library for creating web-based data visualisations. It provides a flexible API to transform data into HTML, SVG, and CSS, making it a robust data visualisation framework. With D3.js, developers can create customizable visualisations and graphics from scratch, including complex multi-variate visualisations that are fully connected and brush-able. It is the most popular library for data-driven documents and can implement everything from simple bar-charts to advanced visualisations.

### 3.5.2 Chart.js

Chart.js is a simple yet very flexible JavaScript library for data visualisation that is popular with web designers and developers. It is a good basic solution for those who do not need many chart types and customization features, but want their charts to look neat, clear and informative at a glance.

### 3.5.3 Bootstrap

Bootstrap is a powerful, extensible, and feature-packed frontend toolkit that improves the design and layout of the web applications. It extends the user experience and enables user-friendly interfaces using prebuilt grid systems and components that reduce development time and the need to create custom stylesheets (cascading style sheets) for each code segment.

### 3.5.4 jQuery

JQuery is a JavaScript library that simplifies the process of adding dynamic content to websites. It facilitates the creation of interactive and engaging user experiences in web-applications. JQuery has, similar to the previously presented tools, an extensive documentation which ensures its ease of use, making it a reliable and efficient framework.

## 3.6 Visualisation methods and types

Definitive methods and types will be determined by the final data structure and schema. Possible methods will be described below and are presented in Figure 10.

### 3.6.1 Line Graphs

A line graph (Figure 10, 1) is used to display quantitative values over a continuous interval or time period, making it a useful tool for showing trends in data over time. When grouped with other lines, individual trends can be compared to one another. While line graphs are useful for displaying patterns in data over time, they are not recommended for displaying discrete data.

### 3.6.2 Node-link Graph

To visualize the connection between different biomaterials a (force-directed) node-link graph (Figure 10, 2) can be employed. A node-link graph (also network diagram, network graph, network map) is a visual representation of interconnected entities using nodes and links. It helps to interpret the network structure and relationships between entities. Nodes are usually circles or dots, and links are represented through lines connecting them. There are two types of network diagrams: undirected and directed. However, they have a limited data capacity and can become unreadable when there are too many nodes.

### 3.6.3 Bar Graphs

A bar chart (Figure 10, 3) uses horizontal or vertical bars to compare numerical data across categories. It shows discrete, categorical data and is different from histograms, which display continuous data. Bar charts are useful for comparing data and identifying patterns.

### 3.6.4 Pie and Donut Charts

Pie and donut charts (Figure 10, 4) show proportional data by dividing a circle into segments. They are useful for comparisons, showing part-to-a-whole, and proportions. However, they have limitations and are not ideal for large datasets or making accurate comparisons between groups.

### 3.6.5 Arc Graphs

Arc diagrams (Figure 10, 5) are a different way of representing a network by placing nodes along a one-dimensional axis and using arcs to display the connections between them. The thickness of the arcs can represent the frequency of connections between nodes, making it useful in identifying co-occurrence patterns. However, arc diagrams do not provide a clear representation of the structure. Additionally, a cluttered diagram with too many links can be challenging to read. Nonetheless, it provides a useful alternative method for investigating and identifying relationships in data. Relationships between authors could be investigated using arc graphs.


**Figure 10.** Overview of visualisation methods

# 4 Biomaterial Marketplace: Use cases, data structures, data formats and entities

## 4.1 Meta use cases, requirements and personas

As identified in the deliverable **D2.1 Knowledge compilation and structural material collection,** there is currently no marketplace focused solely on biomaterials and products made of biomaterials. The few marketplaces available on the web that also include biomaterials have their focus either on generic materials, medical devices, or specific medical applications. Thus, it is one of the main objectives of the project to develop and validate a biomaterial marketplace. To be able to achieve this objective, the meta use cases for the Biomaterial Marketplace have been defined to clearly state the value of the marketplace and which needs it is able to fulfil.

A marketplace, by definition, is able to centralise the exchange of information and goods, facilitating access for the different stakeholders that are interested in them. The same can be said for the Biomaterial Marketplace, which aims to connect suppliers of biomaterials and products made of biomaterials with potential buyers, researchers, policy makers, and other relevant stakeholders. It should also assist the user to identify their need and the solution that may fulfil it.

Based on the information on marketplaces and biomaterials collected for deliverable D2.1 Knowledge compilation and structural material collection, as well as the results from the surveys and interviews performed in WP2 for D2.2 Stakeholder survey, and the knowledge and expertise of the consortium, three meta use cases have been identified that can be seen in Figure 11.

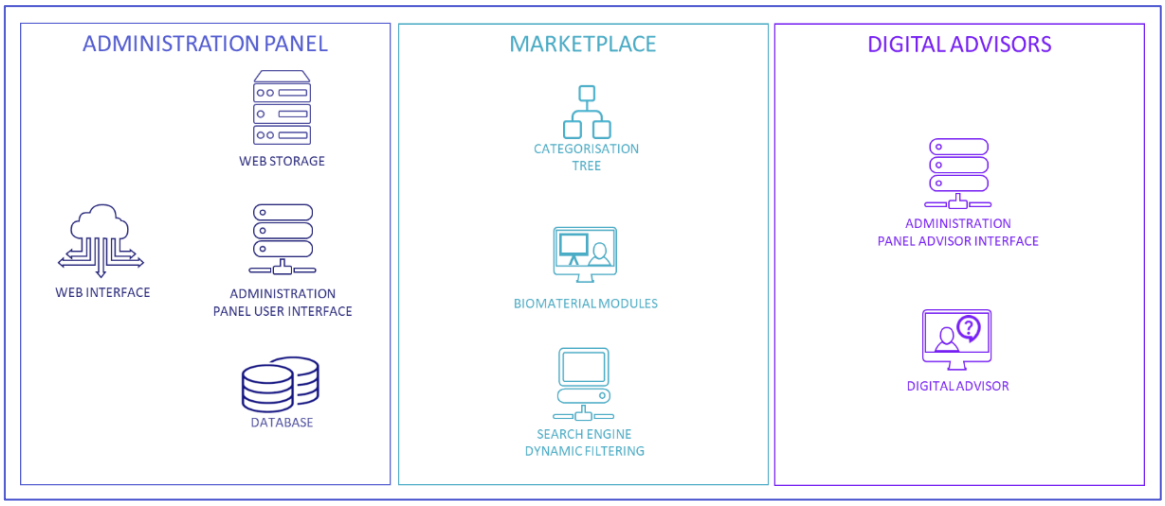



**Figure 11.** Biomaterial Marketplace meta use cases

### 4.1.1 MP-01 Administration panel for the management of supplier's and biomaterial data

#### Use case description

The marketplace system is expected to structure the biomaterial information that is presented to the user, from the products made available by the suppliers to events and news listed in the marketplace. This data will be managed by a central administration panel. The administration panel will be a modern and lightweight application and serve as the core part of the biomaterial marketplace solution.

It will be designed as a single point of entry for all administration user registrations and is used to organise all core database entities (organisations, products, etc.) as well as to manage the user roles and rights. The administration panel will be accessible via **adminbase.biomaterialmarketplace.com** after the Biomaterial Marketplace was launched.

### Marketplace entity structure

In order to add such data, the administrators need to be able to structure the marketplace. They will be able to register and access the administration panel and add the necessary information for the marketplace to be easy to use. These user groups will be responsible for adding a biomaterial specific categorisation tree that will allow products and services to be categorised in the marketplace, add information of necessary supporting entities such as biomaterial related scenarios, events and news and manage the content of the home page, about page and registration page. Administrators will also be responsible for managing the user roles, permissions and other user rights-related data.

### Supplier's profile and biomaterials

The suppliers will use the administration panel to add information about their organisation, biomaterials, products and services. Thus, suppliers will be able to register themselves in the administration panel and will be given rights to add and manage the data that they would like to display for enablers, demanders, researchers, policy makers and other visitors of the marketplace.

The suppliers will be able to add information about their company, add products, assign each product to a biomaterial category, add product information such as properties, attributes, the biomaterial's biocompatibility, product description, and images. Suppliers are also able to add documents to their profiles, which may be biomaterial catalogues, brochures or other relevant documents, and add services and other entities relevant to the biomaterial community.

Figure 12 presents sketches of the layout and structure of the administration panel. It will have a left side navigation bar that lists all the main entities. The administration panel for the suppliers presents only some entities, such as the biomaterials, supplier's profile, documents and services, while the administrator have more entities in order to manage the marketplace data and the entire system.

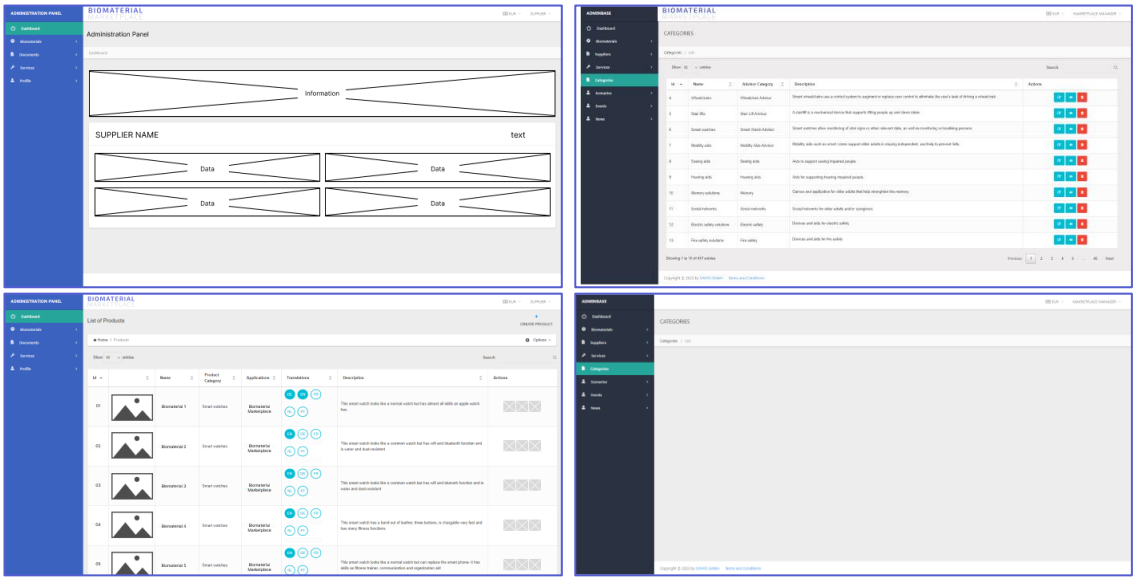



**Figure 12.** Marketplace sketches of MP-01 Administration panel for the management

### Use case specification

**Table 7.** Specification of MP-01 Administration panel for the management

| | | | |
|---|---|---|---|
| **Name of Use Case:** | **Administration panel for the management of supplier's and biomaterial data** | | |
| **Use Case ID:** | MP-01 | | |
| **Created by:** | SYNYO | **Created by:** | SYNYO |
| **Creation date:** | 23.12.2022 | **Creation date:** | 23.12.2022 |
| **Actors:** | Administrator, marketplace content manager, supplier | | |
| **Preconditions:** | User has to be logged in the administration panel as administrator, content manager, or supplier | | |
| **Postconditions:** | Specific content is successfully created, updated or deleted and displayed in the marketplace | | |
| **Results** | The expected result of this use case is that the user has successfully managed the entity the user wants to manage. The system validates the data entered and ensures that all required data for the marketplace have been added. It makes it available in the agreed structure for the marketplace. Thus, removing the chance to break the marketplace system.<br>The validation and notification system assists the user to identify if anything is missing or if something went wrong while performing the action. | | |
| **Evaluation** | The use case will be, in a first iteration, validated and evaluated by members of the BIOMATDB project that will be expected to add content in the marketplace system and access the administration panel to manage the marketplace and add data of a fictional supplier in order to validate the use case. The roles will be either administrator of the system, marketplace content managers or supplier.<br>In a second iteration, suppliers that are stakeholders of the project and have been invited by the BIOMATDB partners to test the administration panel will add their data and biomaterials and validate and evaluate the system and use case. | | |

The table of requirements below presents the main requirements identified at this stage of the project for the creation of an administration panel to manage the biomaterial marketplace.

**Table 8.** Requirements of MP-01 Administration panel

| # | Requirement | Actors |
|---|---|---|
| **MP-01-01** | Register to the administration panel | Administrator, content manager, suppliers |
| **MP-01-02** | Log in the administration panel | Administrator, content manager, suppliers |
| **MP-01-03** | Add biomaterial category tree for products and services | Administrator, content manager |
| **MP-01-04** | Add biomaterial scenarios | Administrator, content manager |
| **MP-01-05** | Link scenarios to categories | Administrator, content manager |
| **MP-01-06** | Add events, news, and other entity data relevant for the marketplace | Administrator, content manager |
| **MP-01-07** | Overview and manage data in the marketplace | Administrator, content manager |
| **MP-01-08** | Manage profile (of supplier) | Suppliers |
| **MP-01-09** | Add products and services, and information about them (including images) | Suppliers |
| **MP-01-10** | Link product and services to categories | Suppliers |
| **MP-01-11** | Add documents, brochures, product catalogues to their supplier profiles and their listed products | Suppliers |

### 4.1.2 MP-02 Biomaterial Marketplace with intelligent search engine

#### Use case description

Marketplaces are used in different industries to provide information on products and services in niche markets including supplier data, product information, usage scenarios, documents, etc. To facilitate the navigation of the user in the marketplace and assist the user to identify and find the best solution for their need, the Biomaterial Marketplace must include processes and system components that assist the user while navigating the marketplace.

The user groups (researchers, enablers, demanders, policy makers, suppliers and any other visitor of the biomaterial marketplace) will be able to use the system to find relevant biomaterials and medical devices, as well as information about their suppliers on the marketplace. Assisted by a complex search engine and a user-friendly interface, the biomaterial marketplace will centrally present the data and interactively support the user in finding the best solution for their needs. The Biomaterial Marketplace will be accessible via **www.biomaterialmarketplace.com** up from a later stage.

**Suppliers**

Suppliers are organisations that provide products or services. The biomaterial marketplace will present the suppliers in a suppliers' list page which gives the user the possibility to search for suppliers using the search engine and various filters. For example, the filter function can include filtering by country, activity types, number of employees, revenues and certifications. Moreover, the supplier's list page can be presented in a list form or in a grid form depending on the preference of the user. Once a certain supplier is found, a detailed page about the supplier is available. In a structured and easy-to-read way, the supplier's detail page displays information such as the supplier's name, activity type, supplier's logo, description, number of employees, revenue, founding year, contact information (location, phone, website) and information about its services or products. Furthermore, a map showing the market's where the supplier is active can be found and a contact form will be available for visitors and users to get in contact with the supplier.

**Biomaterials (products)**

Biomaterials are items that are being displayed for sale. In the case of the biomaterial marketplace, these are biomaterials and products made of biomaterials that have been listed by suppliers. Instead of searching for suppliers, the user will be able to use the search engine and the filter to directly look for biomaterials and certain products made of them. The biomaterials are displayed in a grid or list structure in the biomaterials' list page. The list page allows users to search for biomaterials using the search engine or the filter function. The filter function allows biomaterials to be searched by attributes such as biomaterial categories, biomaterial properties, supplier's name, country and supplier type. After a biomaterial is found, further information about the biomaterial is available in the detail page. The detail page shares information about the biomaterial such as brand names, biomaterial name, description, images of the biomaterial, supplier's name, supplier's location, a contact form to get in contact with the supplier, the biomaterial's attributes, the biomaterial's shop website and external websites where the biomaterial is available. Also, using a smart algorithm, similar biomaterials are displayed below the information of the current biomaterial.

**Categories**

In case the user has little information about the biomaterial, service or its supplier, the user can search the solution by using categories. Categories are classifications of various biomaterials or services. The purpose of categories is to provide an overview of the entire range of biomaterials and services in the catalogue. Categories can be structured in a hierarchical form that resembles a tree. This results in one category possibly having one or more child categories. The end of the tree-structured categorisation are biomaterials and services. Therefore, when searching for biomaterials or services, the user can choose to do so by looking at the different categories. With the support of the search engine, the user can look for a certain category by name or keywords. Once the category is found, detailed information about it is presented to the user, such as category name, an advisor that can support the user to find the perfect biomaterial or service and the category's attributes. Also, a category description including characteristics, types, usage, target groups and further information related to the category is displayed. Finally, the biomaterials and services structured under this category are displayed and related biomaterial categories are also shown.

**Documents (brochures, catalogues)**

Brochures and catalogues are documents that can be downloaded as PDF files and serve as a source of information about biomaterials, services or other supplier related information. These documents are created and managed by suppliers and can be downloaded by biomaterial marketplace users from the supplier's detail page.

**Search bar**

The marketplace will include a search bar that is at all times displayed to the user. The search bar will dynamically suggest results according to the input of the user. The user will be able to either select a solution presented in the result list, which will take the user to a detailed page of the solution, or enter a desired input and view the result related to the input in the form of a list.

**Scenarios**

The scenarios are areas of application of biomaterials that bring information of a specific usage. It supports the classification of biomaterial categories and subsequently biomaterials.

The marketplace will present the user a list of scenarios relevant to biomaterials to assist the user in identifying solutions. The scenarios will present information such as a description, target group, types of solutions and a list of relevant biomaterial categories and relevant biomaterials. The list of relevant biomaterial categories will also have a link to the digital advisors of each category.

The user will be able to either select a biomaterial, which will lead the user to the biomaterial detailed view, or a biomaterial category and/or its digital advisors.

While the first use case of the marketplace is to support and facilitate the search for biomaterials for the user, this use case focuses on informing the user on a set of biomaterial solutions or a specific solution in a user-friendly manner.

The biomaterials information meta use case will present the user with relevant information such as brand names, description about the biomaterial solution, solution's images, attributes, biomaterial category, biocompatibility of the solution, information about the supplier and information about related scenarios and solutions.

The objective is to support the user to gather all the necessary information with as few clicks as possible and efficiently. The use case is split into the following two situations.

**Similarity analysis**

The marketplace will also present the user with the result of an automatic similarity analysis in which similar products, suppliers and/or categories have been identified by the system are listed. This will assist the user to find further relevant solutions and information to its needs. This analysis will be present in the detail pages of biomaterials, and suppliers.

**Filtering**

The filtering will help the user finalise the search for a certain biomaterial solution after the user has already started searching for solutions. The marketplace's wide filtering use case allows the user to funnel the result and eliminate solutions that are not relevant according to different attributes about the solution or its supplier.

The filtering will be a section inside the list view of an entity, such as biomaterial, supplier or service. If the user arrived in the list view by either selecting the entity or using the different workflows of this meta use case, or through the **digital advisors**, the user will be presented with a filter section that allows the user to filter the results even more.

The dynamic filtering section allows the user the ability to filter results according to attributes of the biomaterials, brand names, biocompatibility, geographical restrictions, suppliers, price range and biomaterial category. The filter updates itself also dynamically according to the new results after the user selected a filter option.

Figure 13 presents first sketches of the Biomaterial Marketplace including a list page and a detail page of a biomaterial. The Biomaterial Marketplace will have a top navigation bar with a few buttons on every page, such as a supplier registration button and a top menu bar underneath that allows the user to access the main entities of the marketplace.

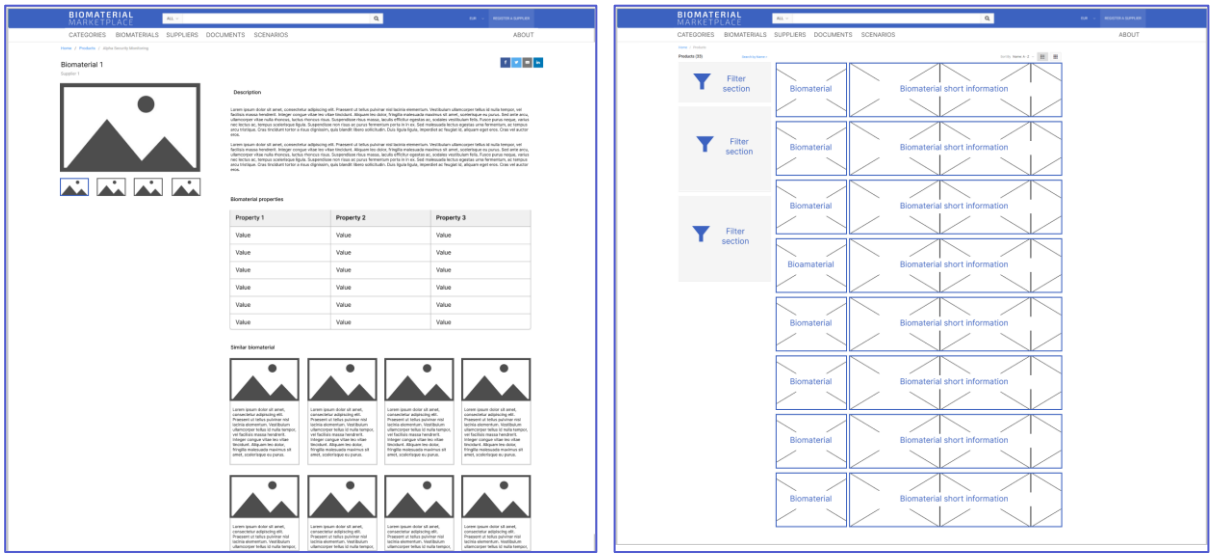



**Figure 13**. Marketplace sketch of MP-02 Biomaterial Marketplace

### Use case specification

**Table 9.** Specification of MP-02 Biomaterial Marketplace

| | | | |
|---|---|---|---|
| **Name of Use Case:** | **Biomaterial marketplace with intelligent search engine** | | |
| **Use Case ID:** | MP-02 | | |
| **Created by:** | SYNYO | **Last update by:** | SYNYO |
| **Creation date:** | 23.12.2022 | **Last revision date:** | 12.01.2023 |
| **Actors:** | Researchers, demanders, enablers, policy makers, suppliers, visitors | | |
| **Preconditions:** | The user has to access the biomaterial marketplace | | |
| **Postconditions:** | The user has either successfully:<br>1. found a biomaterial<br>2. found a biomaterial supplier<br>3. informed themselves about biomaterials and related biomaterial resources | | |
| **Results** | The expected result of this use case is that the user is able to find the desired biomaterial, product or biomaterial supplier, that the user was searching for, and that the user is able to get sufficient information about one or many biomaterials from the marketplace. | | |
| **Evaluation** | The use case will be validated and evaluated in a first iteration by members of the BIOMATDB project that will access the marketplace and search biomaterials using the system search functionalities, and view the information of biomaterials, services, suppliers and other entities available in the marketplace. In a second iteration, other stakeholders outside of the project team will access the marketplace and perform the same action and evaluate the marketplace and use case. | | |

The use case *biomaterial marketplace with intelligent search engine* comprises all requirements related to the marketplace application that is available to any visitor on the web.

The requirements are focused on the search of biomaterials or medical devices, the provision of central biomaterial information in a singular application, as well as the connection between different stakeholders relevant to the biomaterial field (e.g., biomaterial or medical devices suppliers, demanders, enablers, and researchers). These very broad requirements have been broken down into several more concretely defined requirements that can be seen in Table 10.

**Table 10.** Requirements of MP-02 Biomaterial marketplace

| # | Requirement | Actors |
|---|---|---|
| **MP-02-01** | Search for biomaterials or medical devices using a search bar | Enablers, Demanders, Researchers, Visitors, Suppliers |
| **MP-02-02** | Search for suppliers using a search bar | Enablers, Demanders, Researchers, Visitors, Suppliers |
| **MP-02-03** | Search for categories or scenarios using a search bar | Enablers, Demanders, Researchers, Visitors |
| **MP-02-04** | Dynamically update results according to input in search bar | Enablers, Demanders, Researchers, Visitors, Suppliers |
| **MP-02-05** | Search for biomaterials or medical devices by a categories' list and scenarios' list | Enablers, Demanders, Researchers, Visitors, Suppliers |
| **MP-02-06** | View detailed information on biomaterials or medical devices (biomaterial image, description, attributes, biomaterial properties, biocompatibility, supplier, contact information, market availability) | Enablers, Demanders, Researchers, Visitors |
| **MP-02-07** | View detailed information on suppliers (suppliers' image, description, company information, contact information, products, services, market) | Enablers, Demanders, Researchers, Visitors |

| | | |
|---|---|---|
| **MP-02-08** | Contact suppliers through email, contact form, or social media platforms | Enablers, Demanders, Researchers, Visitors |
| **MP-02-09** | View similar solutions | Enablers, Demanders, Researchers, Visitors |
| **MP-02-10** | View similar suppliers or solutions of the same supplier | Enablers, Demanders, Researchers, Visitors |

### 4.1.3 MP-03 Digital advisors

#### Use case description

Intelligent online advisors help different target groups to gather the desired information about products, services, technologies or any other entity they are looking for. The advisor system uses several intelligent query and ranking methodologies in combination with a strong User Experience (UX) to keep the search and selection process very comfortable, especially if the user is fully new to a field.

The digital advisors will support the marketplace users to find a specific biomaterial following a step-by-step process, in which the user answers simple questions about the biomaterial they are searching for. The digital advisors will find biomaterials that fit the answers given by the user. The digital advisors are available by category. This means that each category may have its own digital advisor that assists users to find the best biomaterial for their needs from the category.

The digital advisor uses category specific questions to guide the user to the solutions relevant to the user's needs. Simple questions such as price range, compatibility of biomaterials and other binary ("yes" or "no") questions about biomaterial attributes are asked to the user. After this set of questions is answered, the digital advisor redirects the user to a list page of solutions that fit the answers of the user.

Figure 14 is a first sketch of the design and layout of the digital advisor. Specifically, it displays one step in the advisor process. As can be seen, it asks the user a question and gives them clear answer options. The blue progress bar on the bottom displays the user's progress. It thereby informs the users how many of the questions they already answered.

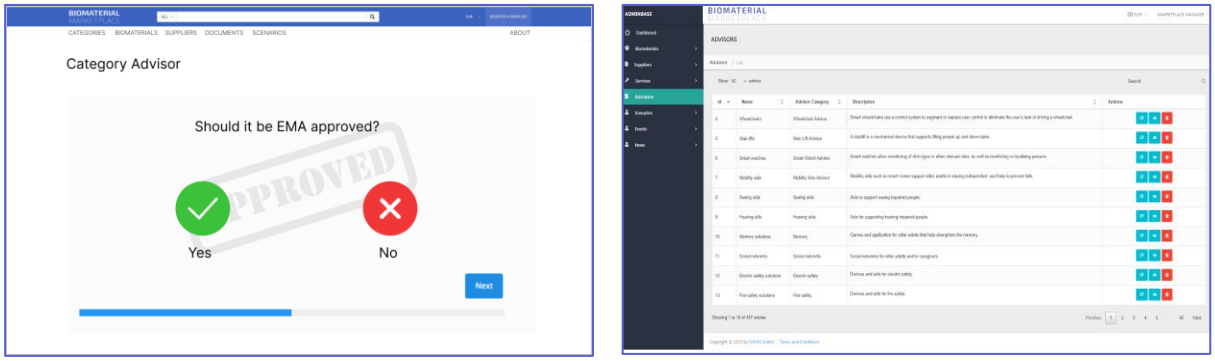



**Figure 14.** Marketplace sketch of MP-03 Digital advisor

This use case is part of both the administration panel and the marketplace systems. It expects administrators and marketplace content managers to create advisor steps, questions, and answer options for each category. Thus, a digital advisor component will be available in the administration panel for the management of the advisors. On the other hand, the functionality of the digital advisor to support the user will be available in the marketplace in connection to the available categories. The feature will be accessible via **advisors.biomaterialmarketplace.com** later on.

### Use case specification

**Table 11.** Specification of MP-03 Digital advisor

| | | | |
|---|---|---|---|
| **Name of Use Case:** | **Digital advisors** | | |
| **Use Case ID:** | MP-03 | | |
| **Created by:** | SYNYO | **Created by:** | SYNYO |
| **Creation date:** | 23.12.2022 | **Creation date:** | 23.12.2022 |
| **Actors:** | Researchers, Demanders, Enablers, Visitors | | |
| **Preconditions:** | User has accessed the marketplace and is in the scenarios page or categories page. | | |
| **Postconditions:** | The user found one or many biomaterials according to the user's needs. | | |
| **Results** | The result expected from this use case is a list of biomaterials that fit exactly the needs of the user, according to the user's answers. | | |
| **Evaluation** | The use case will be validated and evaluated in a first iteration by members of the BIOMATDB project that will access the marketplace and use the digital advisors to find biomaterials. The validation and evaluation | | |

| | |
|---|---|
| | will focus on the correctness of the resulting biomaterial according to the answers, and on the relevance of the questions and the answer options. In a second iteration, other stakeholders outside of the project team will access the marketplace and be able to validate and evaluate the system and the use case. |

The requirements focus on making the advisor easy to use for the user and be of substantial benefit to the procurement of biomaterials or medical devices when using the biomaterial marketplace. The user groups involved in this use case are biomaterial or medical device enablers, demanders, researchers, and visitors, as can be seen in Table 12.

**Table 12.** Requirements of M-05 Digital advisors

| # | Requirement | Actors |
|---|---|---|
| **MP-03-01** | Be able to access digital advisors for each category in order to find products that fit the needs | Enablers, Demanders, Researchers, Visitors |
| **MP-03-02** | Perform a step-by-step process to find the best solutions to the user's needs | Enablers, Demanders, Researchers, Visitors |
| **MP-03-03** | Answer simple questions to filter the solutions that are supported by user friendly icons and user interface resources | Enablers, Demanders, Researchers, Visitors |
| **MP-03-04** | Get a list with solution(s) according to the answers given | Enablers, Demanders, Researchers, Visitors |

### 4.1.4 Personas


**Figure 15.** Marketplace Persona: Biochemist

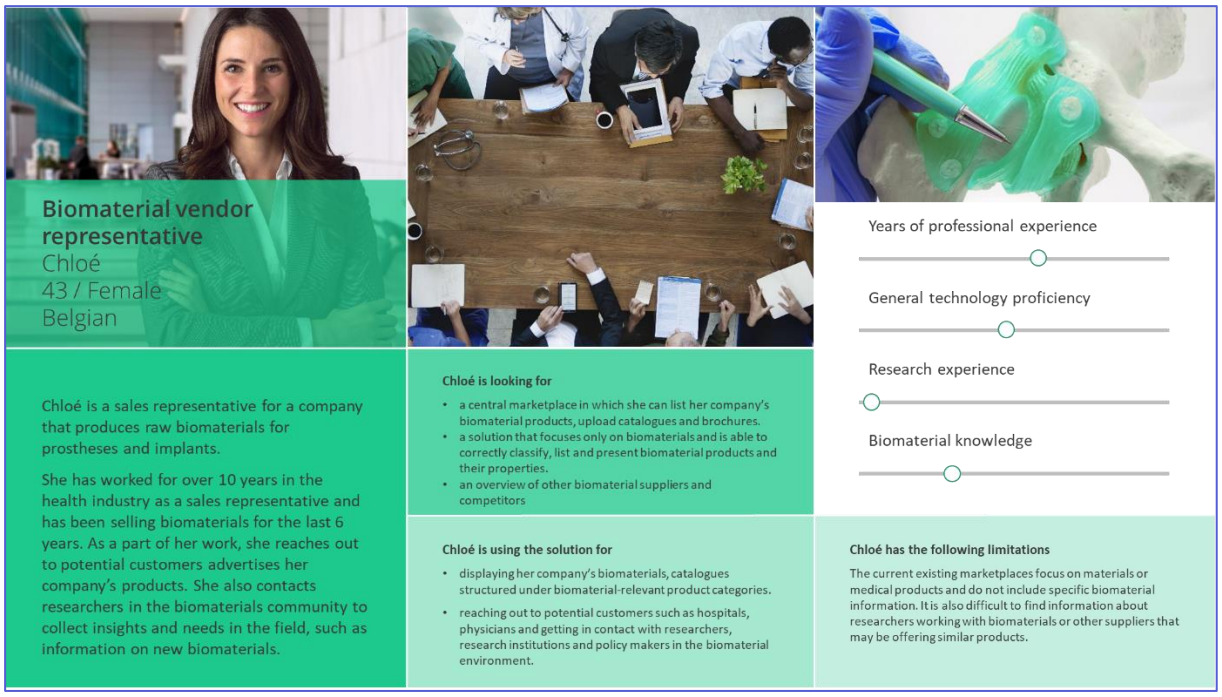



**Figure 16.** Marketplace persona: Biomaterial vendor representative


**Figure 17.** Marketplace persona: Hospital purchasing agent


**Figure 18.** Marketplace persona: Local government councillor

## 4.2 Data sources and data flow

### 4.2.1 Data source: Biomaterial and medical device suppliers

The data for the Biomaterial Marketplace comes mainly from suppliers that register on the marketplace and add information on their company, products, services and other entities to the marketplace.

To reach as many suppliers as possible while still being accurate on the type of supplier that is being reached, the BIOMATDB project team will follow certain methods:

1. Contact to suppliers directly through the BIOMATDB partners, especially, partners active in the industry and with numerous contacts to biomaterial and medical device suppliers.
2. Collection of suppliers information from open data sources via the Biomaterial Database. This may include datasets directly on biomaterials, but also datasets on publications and articles that include names of suppliers and that the database is able to categorise as suppliers.
3. Through marketing and project dissemination on social media, the project website, events and fairs.

The suppliers will all be invited, and will add their information themselves to the marketplace.

### 4.2.2 Data source: EU portals

While the data for the core modules (suppliers, products, services etc.) is all entered manually, the Biomaterial Marketplace also foresees multiple extension modules. The extension modules are EU projects, EU project partners, EU funding and EU tenders. Because of the EU context, these extension modules use the following external data sources provided by official EU institutions:

- European Data Portal
- CORDIS
- EC Funding & Tenders Portal
- TED

### 4.2.3 Data flow

The data flow within the marketplace system is a crucial aspect that drives its functionality. Irrespective of the data source, all information undergoes a unified data flow within the technical system. Once received, the data is processed, stored, and prepared for accessibility through an API endpoint.

Upon completion of the data processing, the prepared information is made available for consumption on the application frontend. The application directly retrieves the required data from the API endpoints, enabling seamless presentation of the information to users. This direct integration ensures that the frontend displays the most up-to-date and accurate data retrieved from the underlying technical system. An overview of the system's data flow can be seen in Figure 19.

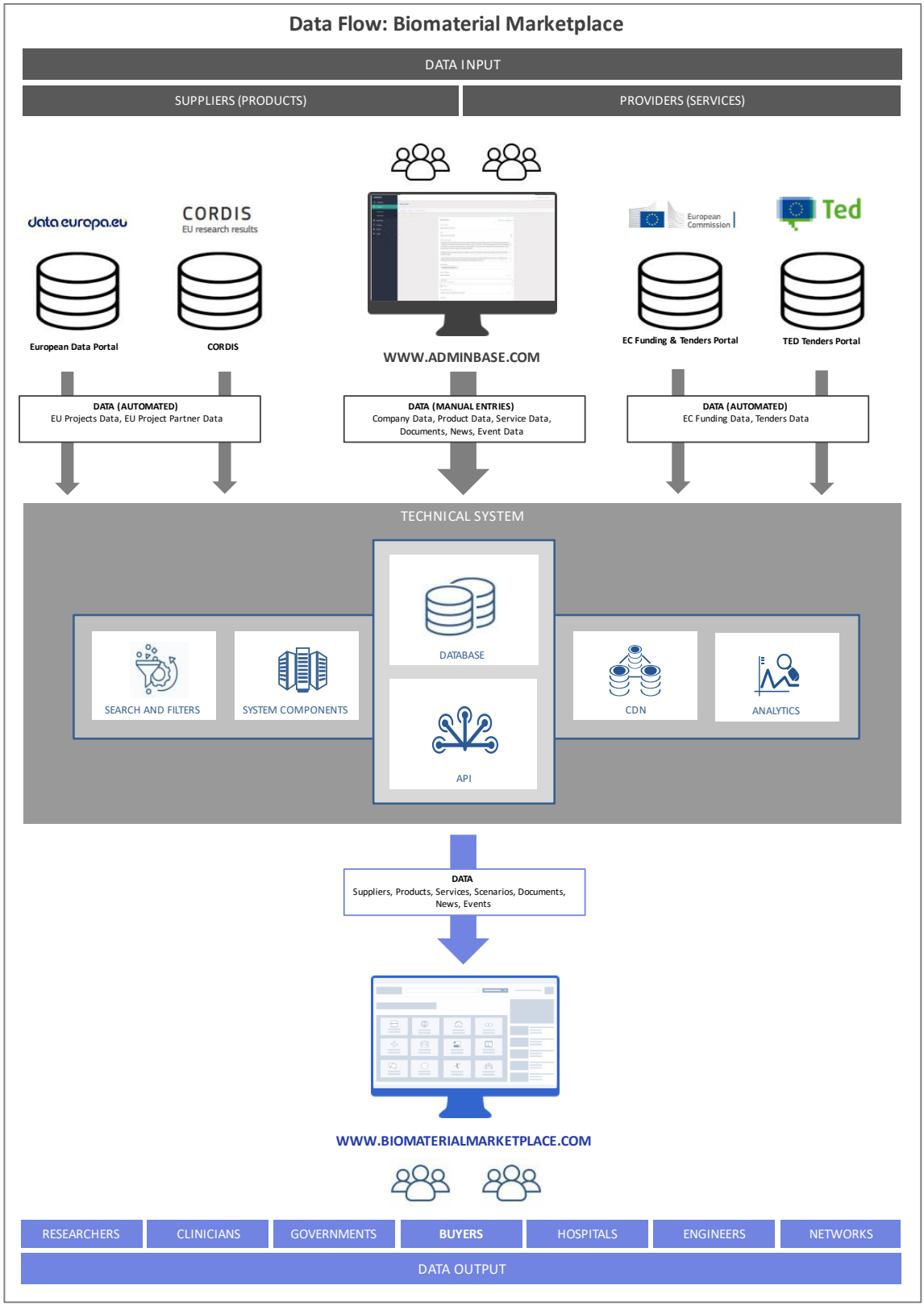



**Figure 19**. Data Flow Biomaterial Marketplace

## 4.3 Data entities, format and types

### 4.3.1 Data Entities (Core Modules)

**Table 13.** Data Entities (Core Modules)

| Name | Description | Data Formats |
|---|---|---|
| **Suppliers** | Suppliers are manufacturers, developers and/or vendors of products or services, in the context of biomaterials. This group includes companies that develop, manufacture and/or sell raw biomaterials or biomaterial-based products | String, Numeric, Images |
| **Products** | In the context of the Biomaterial Marketplace, all products are related to the field of biomaterials. Products can be medical devices (made of, or containing biomaterials), therapies based on biomaterials such as ATMP, components of medical devices, raw or shaped biomaterials and others. | String, Numeric, Images |
| **Services** | Services are similar to products and are also offered by suppliers. Like products, services present short information about their supplier and more in detail information about the service being offered and their characteristics. | String, Numeric, Images |
| **Scenarios** | Scenarios are use cases of biomaterials that bring information of a specific usage. It supports the classification of biomaterial categories and subsequently biomaterials. | String, Numeric, Images |
| **Documents** | Documents are files that can be downloaded as PDF and serve as a source of information about products, services or other supplier related information. | String, Numeric, PDF |
| **News** | News aim to provide up-to-date information related to biomaterials. This includes news articles, press releases or announcements from suppliers. | String, Numeric, Images |
| **Events** | Events are biomaterial-related meetings and activities such as conferences, webinars, workshops and seminars. | String, Numeric, Date/Time, Images |

### 4.3.2 Data Entities (Extension Modules)

**Table 14.** Data Entities (Extension Modules)

| Name | Description | Data Formats |
|---|---|---|
| **EU Projects** | EU projects are projects that are financed or co-financed by the European Union through its various funding programs. They aim to promote research, innovation and development in various fields. | CSV, XML |
| **EU Project Partners** | EU Project Partners are organisations that participated or are currently participating in EU-funded projects. Each partner has a role within a project (coordinator or participant) and an organisation type. | CSV/XML |
| **EU Funding** | EU Funding refers to the announcements made by the European Union inviting organisations to submit proposals for funding opportunities in various areas. | REST/JSON |
| **EU Tenders** | EU tenders are procurement opportunities for goods, services or works that are issued by the European Union. | XML |

### 4.3.3 Data Examples (Core Modules)

All rights of the following examples, including trademarks for specific products, belong to the referenced manufacturers.

#### Suppliers

The aim of a supplier module is to allow suppliers to register and showcase their organisation and products (and other solutions) to potential customers. Suppliers are manufacturers, developers and/or vendors of products or services, in the context of biomaterials. This group includes companies that develop, manufacture and/or sell raw biomaterials or biomaterial-based products like medical devices.

Showcases for existing biomaterial suppliers are Medtronic, Orchid Orthopedic Solutions, ARTIQO, INNOTERE and Sagemax.

**Table 15.** References Suppliers

| Supplier | URL |
|---|---|
| **Medtronic** | https://www.medtronic.com/us-en/index.html |
| **Orchid Orthopedic Solutions** | https://www.orchid-ortho.com/ |
| **ARTIQO** | https://artiqo.de/en/ |
| **INNOTERE** | https://www.innotere.de/ |
| **Sagemax** | https://sagemax.com/ |

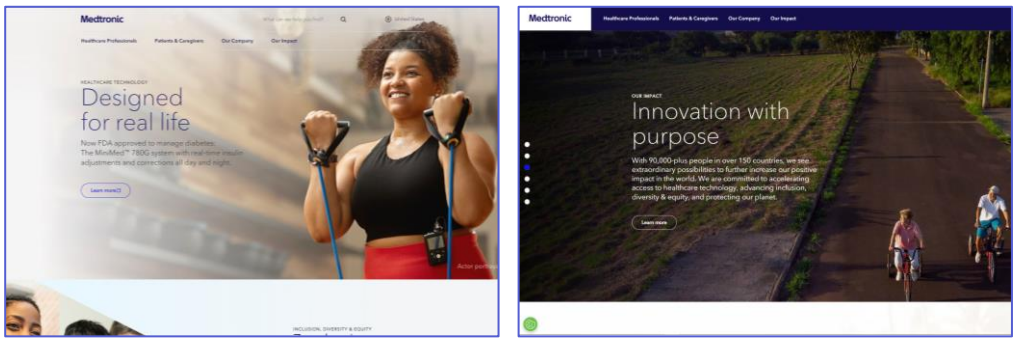



**Figure 20.** Supplier Example (1) Medtronic

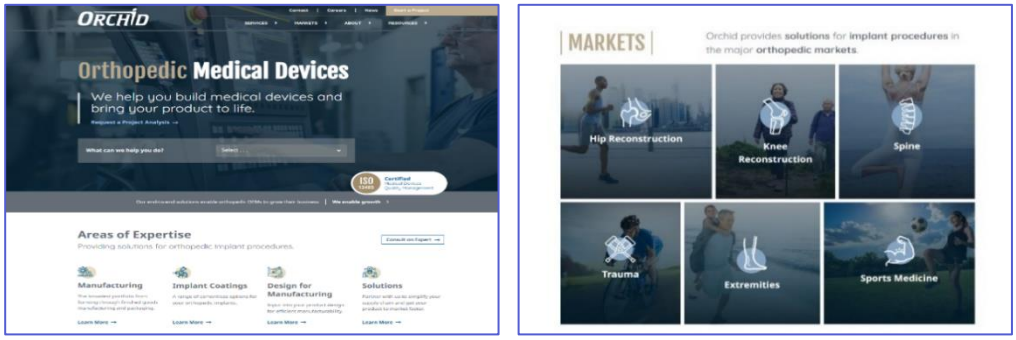



**Figure 21.** Supplier Example (2) Orchid Orthopedic Solutions

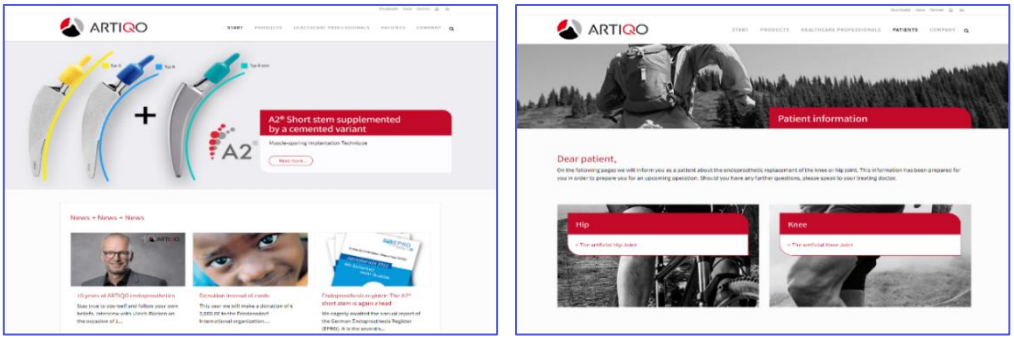



**Figure 22.** Supplier Example (3) ARTIQO

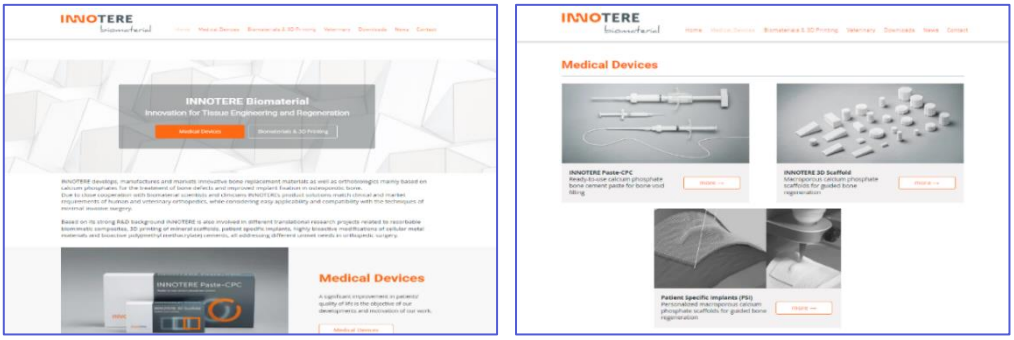



**Figure 23.** Supplier Example (4) INNOTERE

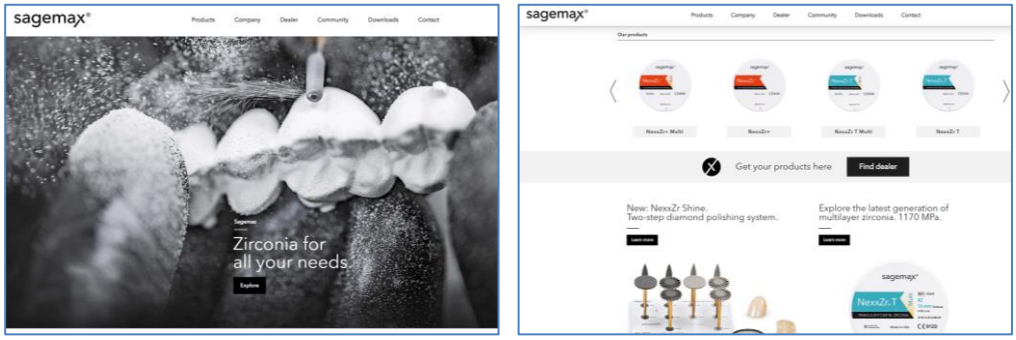



**Figure 24.** Supplier Example (5) Sagemax

### Products

The Biomaterial Marketplace provides a comprehensive listing of all available biomaterial products offered by the registered suppliers. This module helps potential buyers to easily search for and filter the products by various attributes. Products can be medical devices (made of, or containing biomaterials), therapies based on biomaterials such as ATMPs, components of medical devices, raw or shaped biomaterials, and others.

Examples for Biomaterial Products are the A2 Short Stem by ARTIQO, the Impella RP heart pump, the ActivaPin by Bioretec and the Inion BioRestore.

**Table 16.** References Products

| Product | URL |
|---|---|
| **A2 Short Stem** | https://artiqo.de/en/products/hip-stem/a2-short-stem/ |
| **Impella RP heart pump** | https://www.heartrecovery.eu/products-and-services/impella/impella-rp |
| **ActivaPin** | https://bioretec.com/products/1/activapin-bioabsorbable-pin |
| **Inion BioRestore** | https://www.inion.com/product/inion-biorestore-bone-graft-substitute/ |

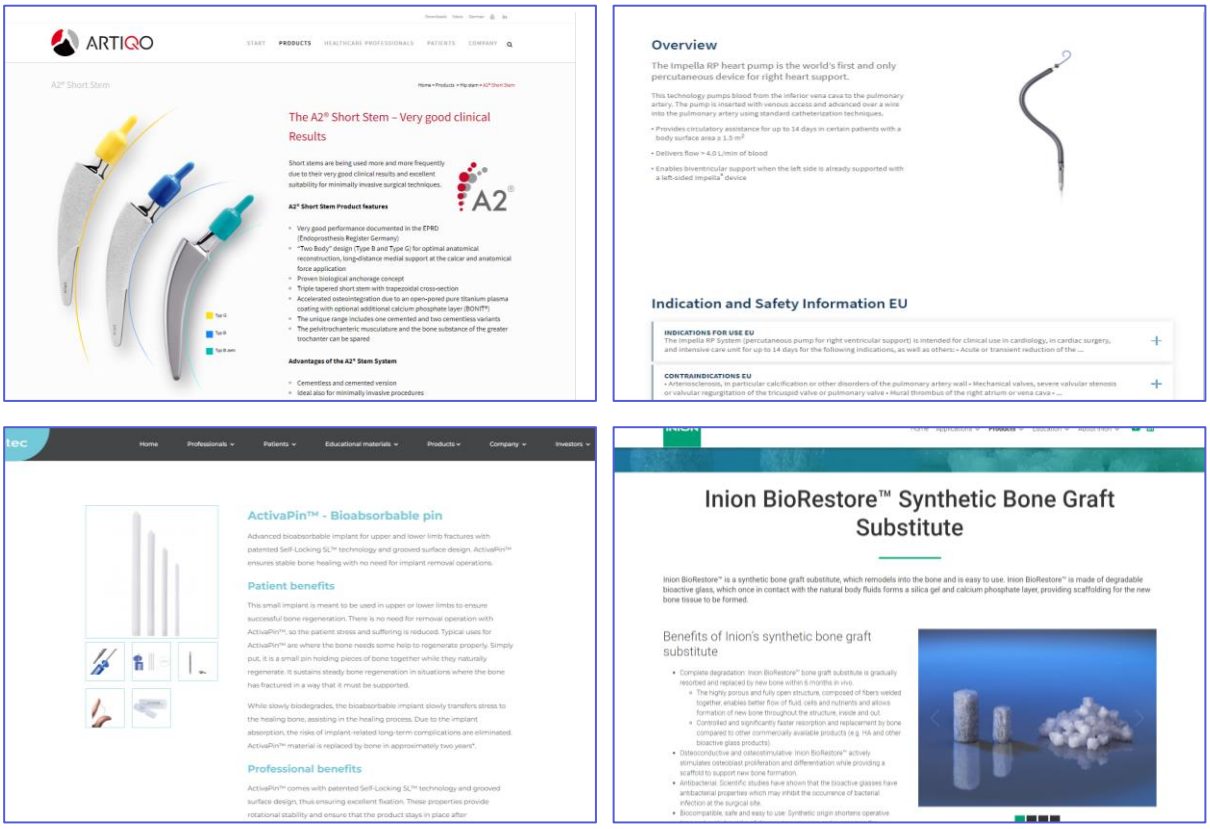



**Figure 25.** Product Examples

## Services

Among products, the Biomaterial Marketplace also lists services related to biomaterials that are offered by suppliers. The service module allows suppliers to showcase their expertise and provide a wider range of offerings to customers beyond just selling physical products.

Some existing suppliers offering biomaterial-related services are Eurofins, Biofarma, Biochem and Abich.

**Table 17.** References Services

| Services | URL |
|---|---|
| **Eurofins Services** | https://www.eurofins.com/our-services/ |
| **Biofarma Group Services** | https://www.biofarmagroup.it/en/services/ |
| **Biochem Services** | https://www.biochem-bcm.com/en/servizi/ |
| **Abich Services** | https://www.abich.it/en/ |

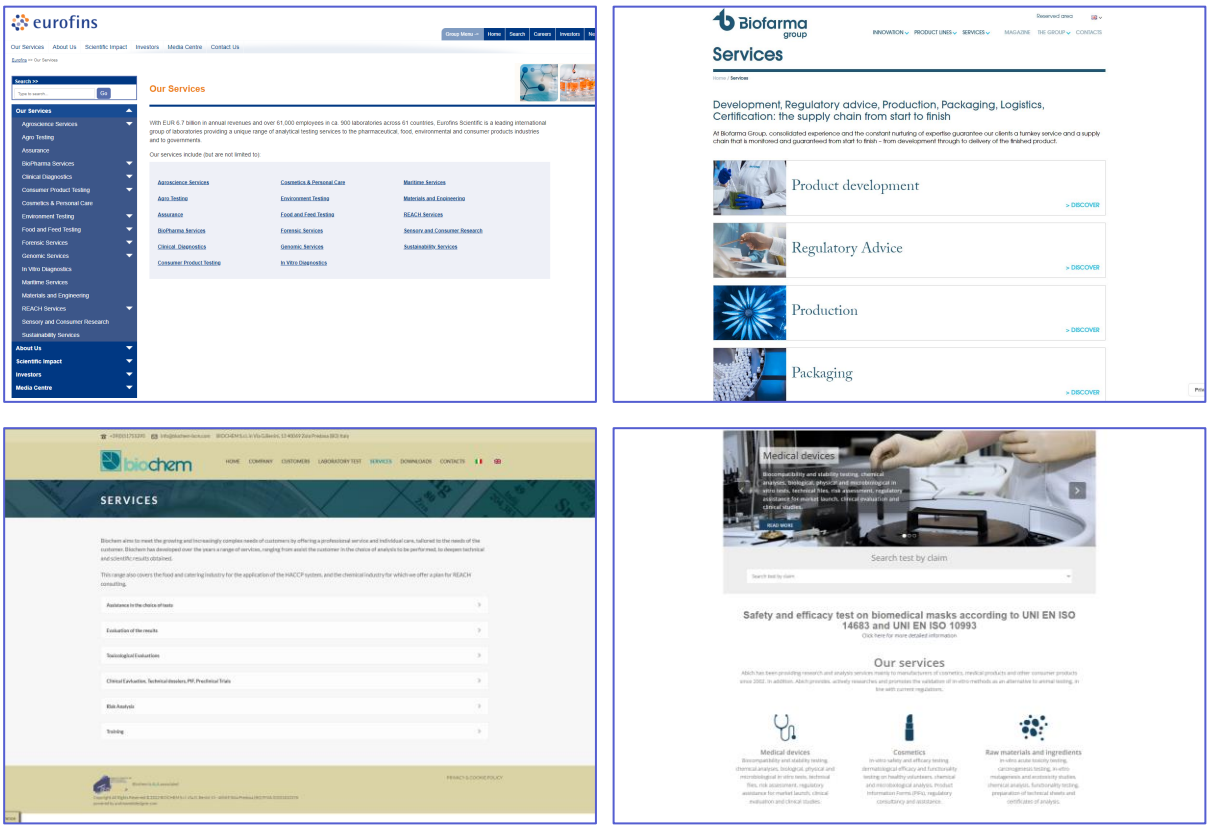



**Figure 26.** Service Examples

### Scenarios

The marketplace will present the user a list of scenarios relevant to biomaterials to assist the user in identifying solutions. The scenarios will present information such as a description, target group, types of solutions and a list of relevant biomaterial categories and relevant biomaterials.

The supplier MED-EL provides a good example of what these scenarios can look like. MED-EL offers hearing implants for people suffering from hearing loss. As there are mainly four different types of hearing loss, each to be treated differently, MED-EL is implementing hearing solutions for these four causes ('scenarios'). As it can be seen in the figure below, each scenario includes a short description, followed by solutions (in this case products) for the respective situation.

**Table 18.** References Scenarios

| Scenarios | URL |
|---|---|
| **MED-EL Types of Hearing Loss** | https://www.medel.com/about-hearing/types-of-hearing-loss |

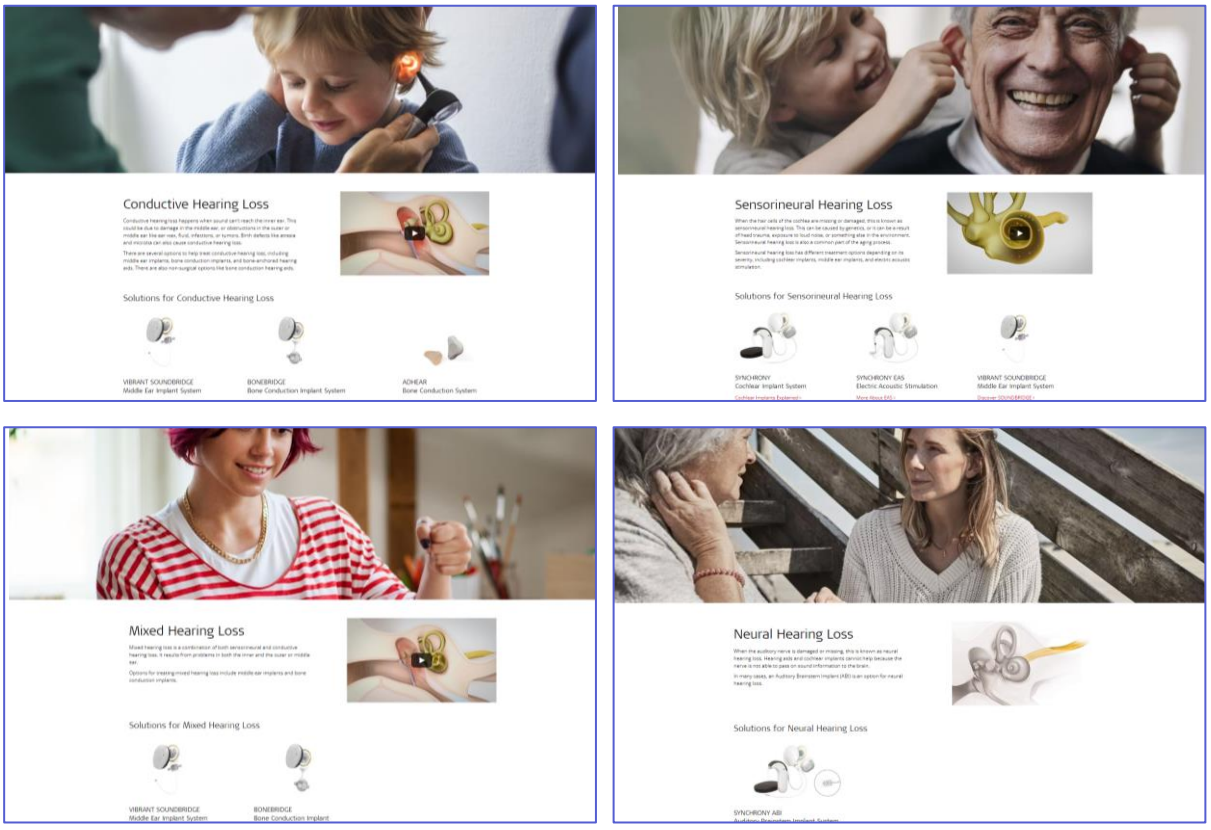



**Figure 27.** Scenarios Example MED-EL

### Documents

The Biomaterial Marketplace includes a location for all relevant documents related to the biomaterial products and their suppliers. With the document module, suppliers are able to upload documents such as product brochures, manuals, technical data sheets or safety information about biomaterials and related products.

Some biomaterial suppliers, like Exatech and Neobiotech, offer an own module on their web page as central point of contact for relevant documents, for example brochures or documentations as it can be seen in the figures below the reference table.

**Table 19.** References Documents

| Documents | URL |
|---|---|
| **Exatech Resource Library** | https://www.exac.com/resource-library/ |
| **Neobiotech Product Catalogues** | https://www.neobiotech.com/product-detail/is-ii-active |

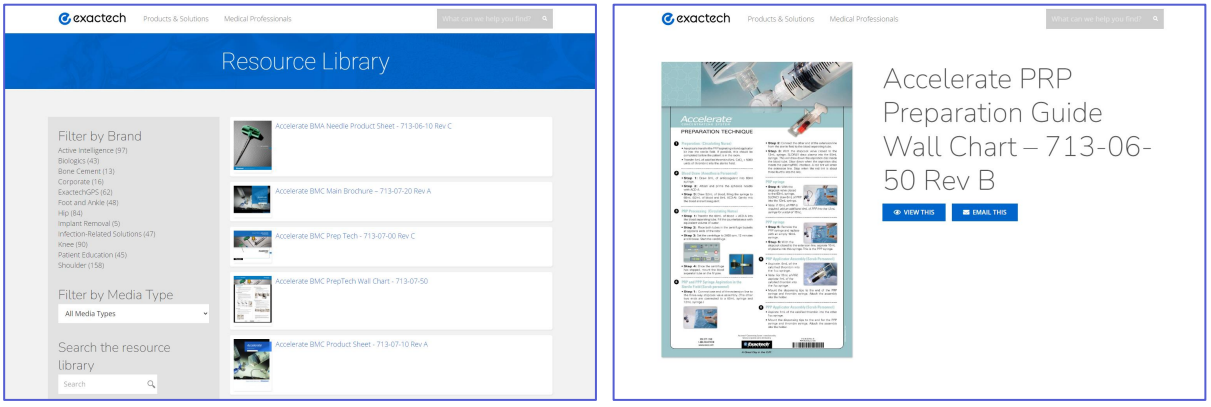



**Figure 28.** Document Example (1) Exactech Resource Library

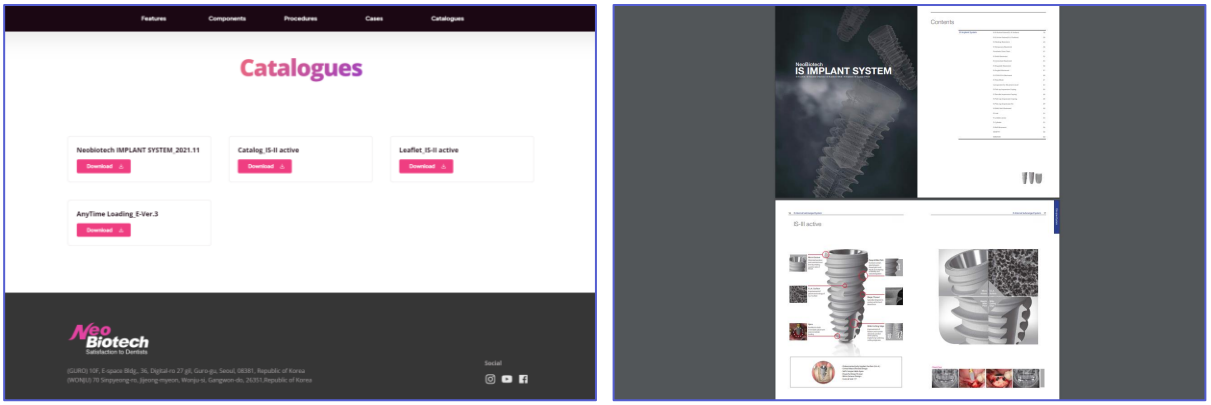



**Figure 29.** Document Example (2) Neobiotech Product Catalogues

### News

The news module on the Biomaterial Marketplace aims to provide up-to-date information and news related to biomaterials. This includes news articles, press releases or announcements from suppliers. The purpose is to keep the users informed about the latest developments and innovations in the biomaterials field.

There are various biomaterial suppliers that keep their audience up-to-date on their website in form of an own news page or something similar – for example Alcon, Anika, Bonalive and CeramTec.

**Table 20.** References News

| News | URL |
|---|---|
| **Alcon Newsroom** | https://www.alcon.com/newsroom |
| **Anika News** | https://anika.com/about-anika/news/ |
| **Bonalive News** | https://www.bonalive.com/en/news/ |
| **CeramTec News** | https://www.ceramtec-group.com/en/news-events/news-overview |

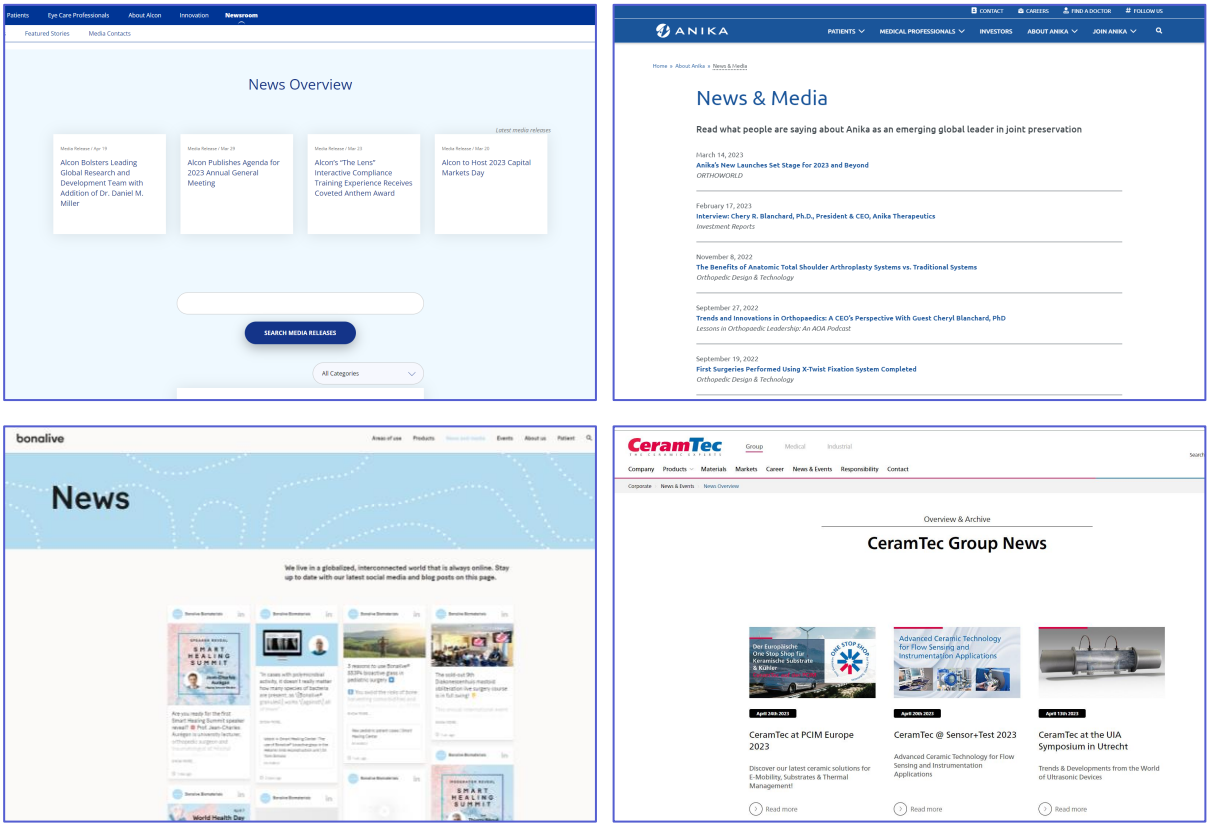



**Figure 30.** News Examples

### Events

The main objective of the event module is the promotion of biomaterial-related events such as conferences, webinars, workshops and seminars. The module allows organisers to create their events with relevant information, including a title, the date(s), a description and the type of the event. In the end, the event module on the Biomaterial Marketplace should serve as a centralised location for the biomaterial community to discover and participate in events related to their industry.

Suppliers with an event module on their web appearance are Bonalive, Bioteck, 3-D Matrix and Coloplast.

**Table 21.** References Events

| Events | URL |
|---|---|
| **Bonalive Events** | https://www.bonalive.com/en/events/ |
| **Bioteck Events and Fairs** | https://bioteck.com/en/events-and-fairs/ |
| **3-D Matrix Events** | https://3dmatrix.com/about/events/ |
| **Coloplast Financial Calendar & Events** | https://investor.coloplast.com/investor-relations/financial-calendar-and-events/ |

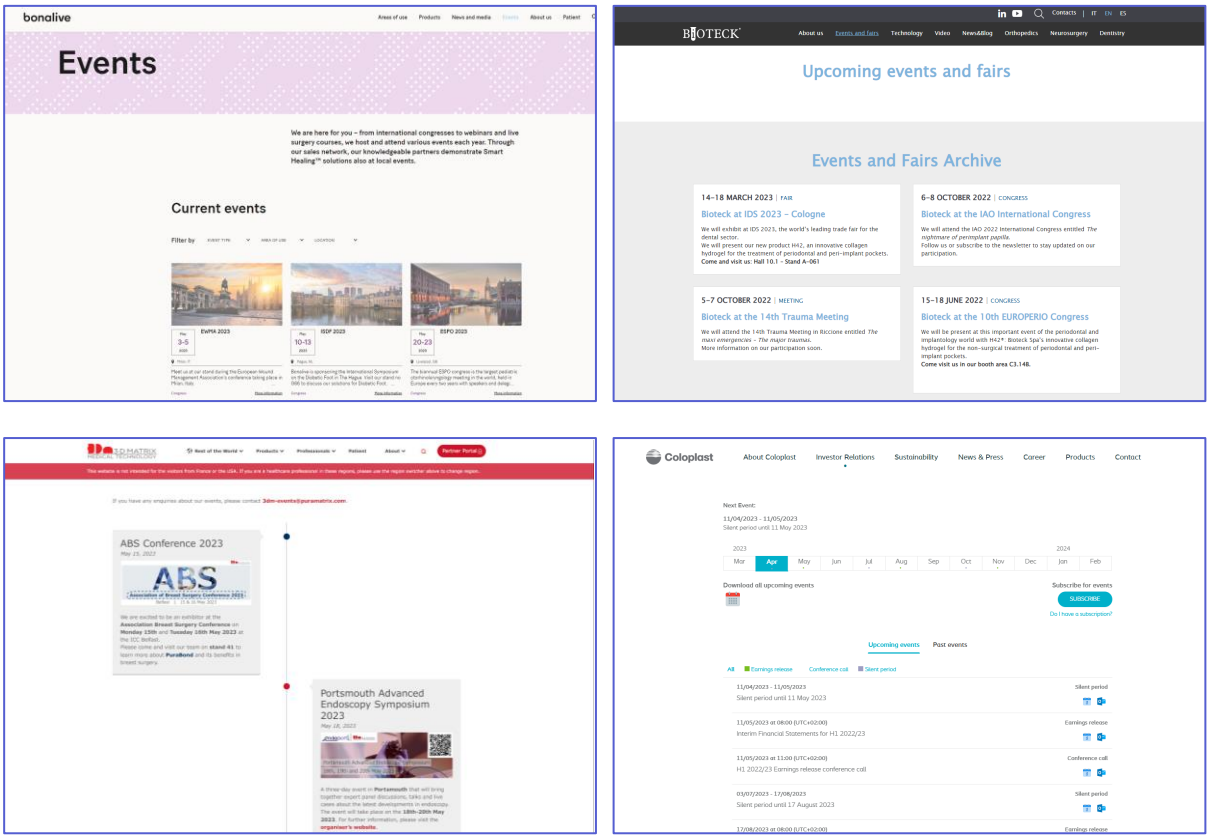



**Figure 31.** Event Examples

### 4.3.4 Data Examples (Extension Modules)

#### EU Projects

EU projects are projects that are financed or co-financed by the European Union through its various funding programmes. They aim to promote research, innovation and development in various fields such as science, technology, energy, environment and social sciences. The EU provides funding opportunities to support innovative ideas and solutions that address the societal challenges facing Europe. These projects typically involve collaborations between organisations from different countries and sectors.

CORDIS (Community Research and Development Information Service) is the European Union's primary information service for research and development projects funded by the EU. It provides a comprehensive database of EU-funded projects and their results, as well as news and information on EU research and innovation policy.

**Table 22.** References EU Projects

| Description | URL |
|---|---|
| **Project List** | https://cordis.europa.eu/search?q=%27biomatdb%27&p=1 |
| **Project Detail View** | https://cordis.europa.eu/project/id/101058779 |

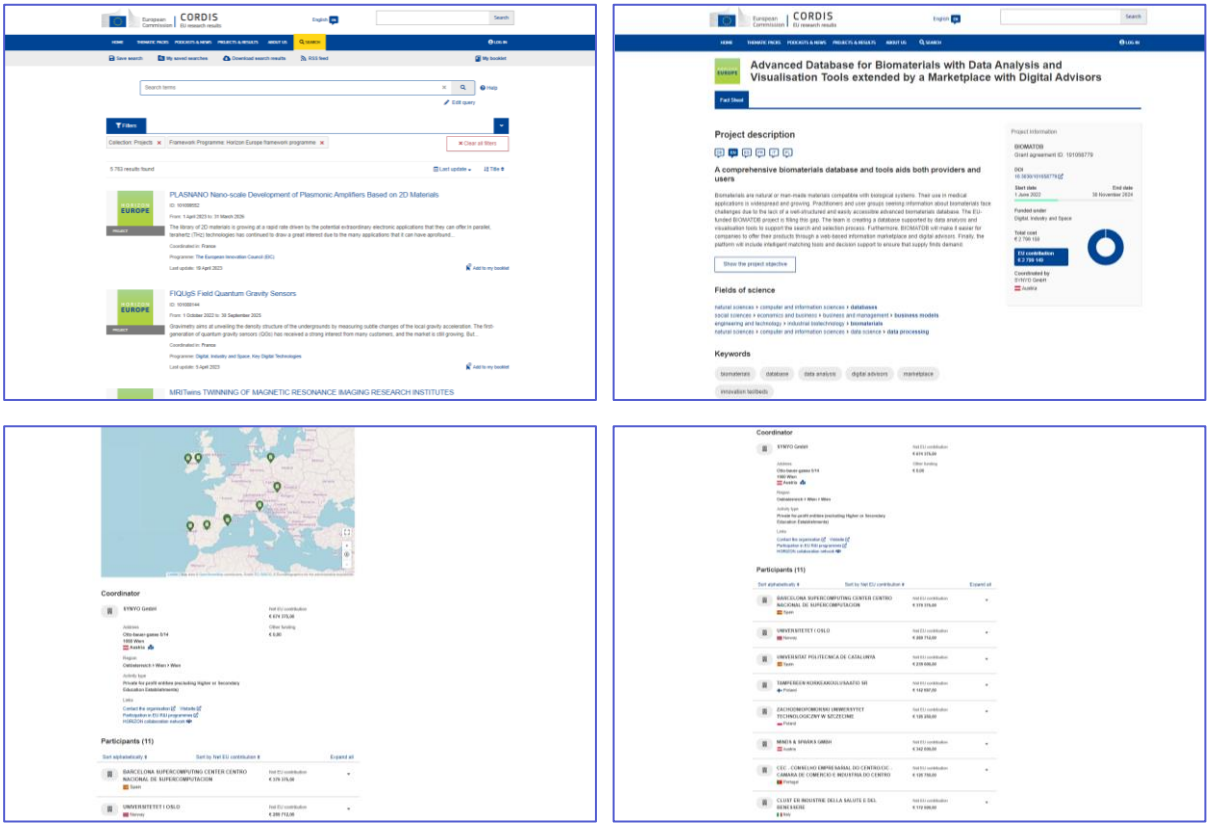



**Figure 32.** EU Projects Example (CORDIS)

### EU Project Partners

EU Project Partners are organisations that participated or are currently participating in EU-funded projects. Each partner has a role within a project (coordinator or participant) and an organisation type, such as research organisation or SME (Small or medium-size enterprise). Furthermore, they provide descriptions and keywords about their application fields and expertise, as well as information about their project participations.

**Table 23.** References EU Project Partners

| Description | URL |
|---|---|
| **Partner List** | ec.europa.eu * |
| **Partner Detail View** | ec.europa.eu * |

* abbreviated URL

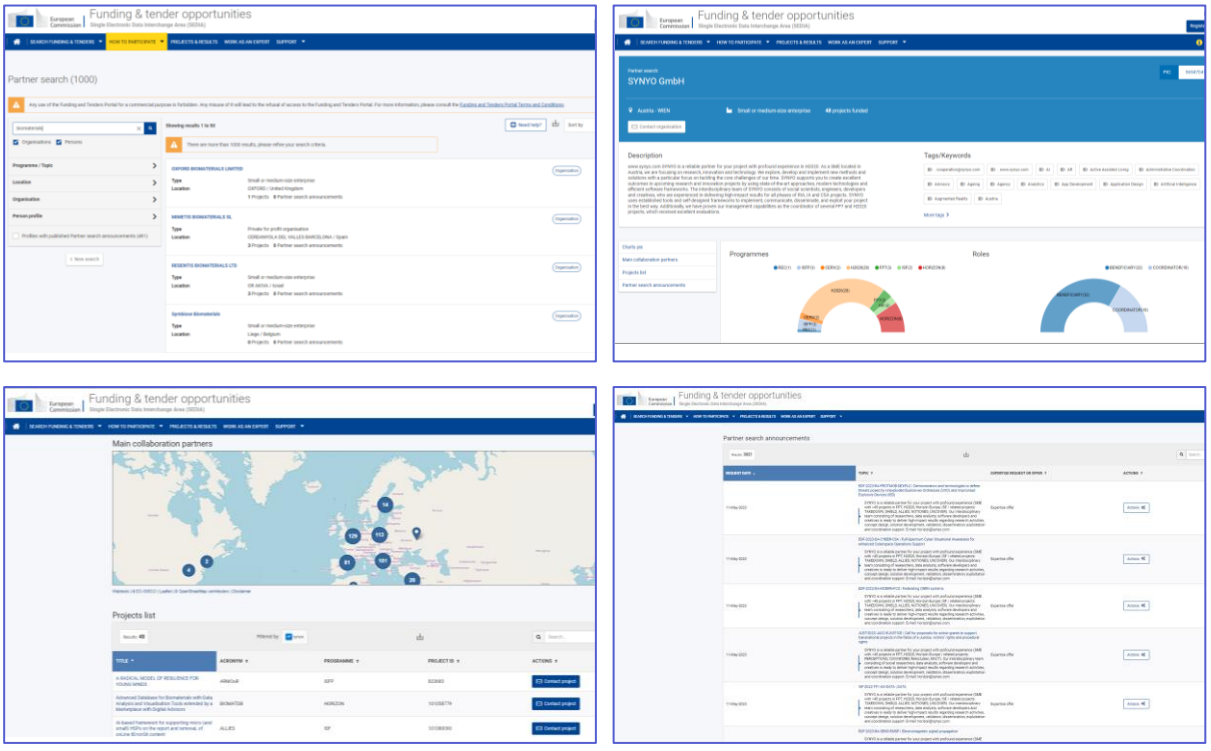



**Figure 33.** EU Project Partners Example (EC Funding & Tender Portal)

### EU Funding

EU Funding refers to the announcements made by the European Union (EU) inviting organisations to submit proposals for funding opportunities in various areas. These calls provide details on the specific topics, eligibility criteria, funding amounts, deadlines and evaluation process. Organisations can apply for funding by submitting proposals that align with the objectives of the call and meet the necessary requirements.

The EC Funding & Tender Portal is an online platform provided by the European Commission that serves as a central access point for different types of funding. The portal provides a user-friendly interface for searching and applying for funding opportunities, as well as managing the application process and monitoring project progress.

**Table 24.** References EU Funding

| Description | URL |
|---|---|
| **Project Funding List** | ec.europa.eu * |
| **Project Funding Detail View** | ec.europa.eu * |

* abbreviated URL

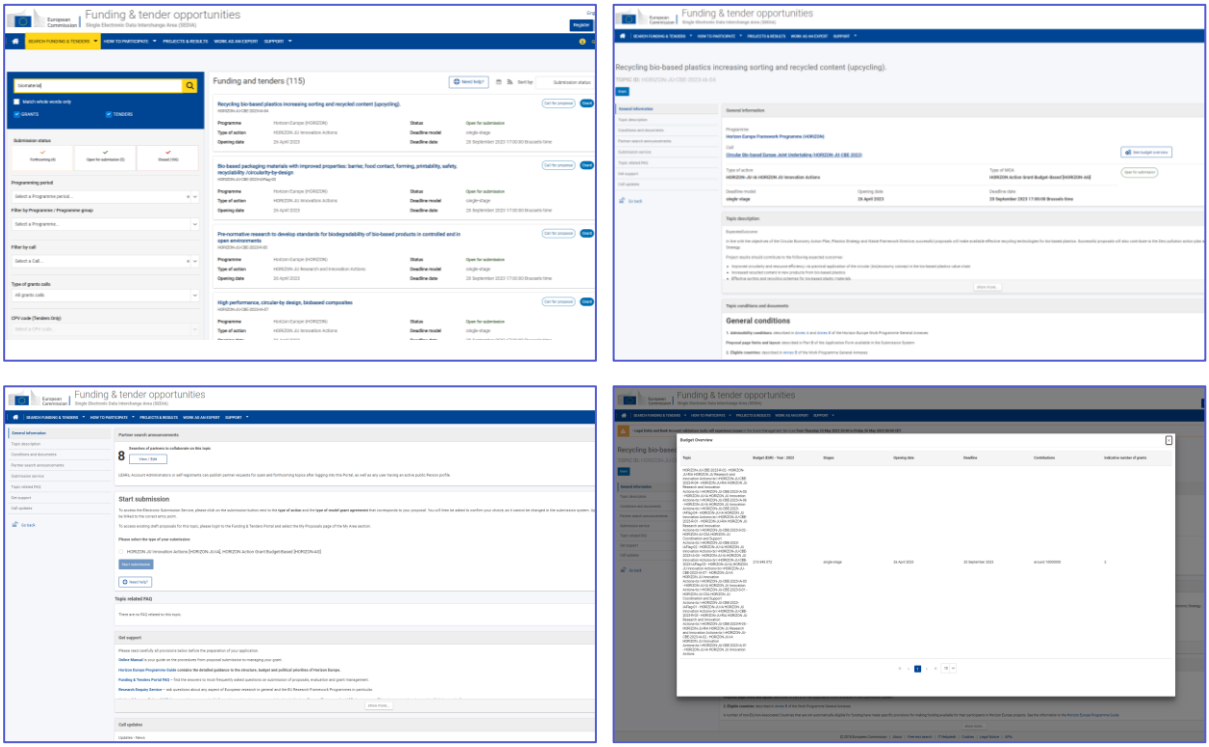



**Figure 34.** EU Funding Example (EC Funding & Tender Portal)

### EU Tenders

EU tenders are procurement opportunities for goods, services or works that are issued by the European Union. Tenders aim to promote fair competition, transparency and equal access to public procurement contracts. EU tenders cover a wide range of sectors, like construction, IT, research and development and consulting. Organisations interested in bidding for EU tenders must comply with specific eligibility and technical requirements.

TED (Tenders Electronic Daily) is an online portal that provides access to public procurement notices from all EU member states, as well as Iceland, Liechtenstein, Norway and Switzerland. TED is the primary source of information for businesses, organisations and individuals interested in participating in public procurement opportunities. It enables users to search and filter procurement notices based on various criteria, such as type of contract, location and deadline.

**Table 25.** References EU Tenders

| Description | URL |
|---|---|
| **Tender List** | https://ted.europa.eu/TED/search/search.do * |
| **Tender Detail View** | https://ted.europa.eu/udl?uri=TED:NOTICE:254311-2023:TEXT:EN:HTML&src=0 |

* Search results not shareable via URL; to be submitted manually

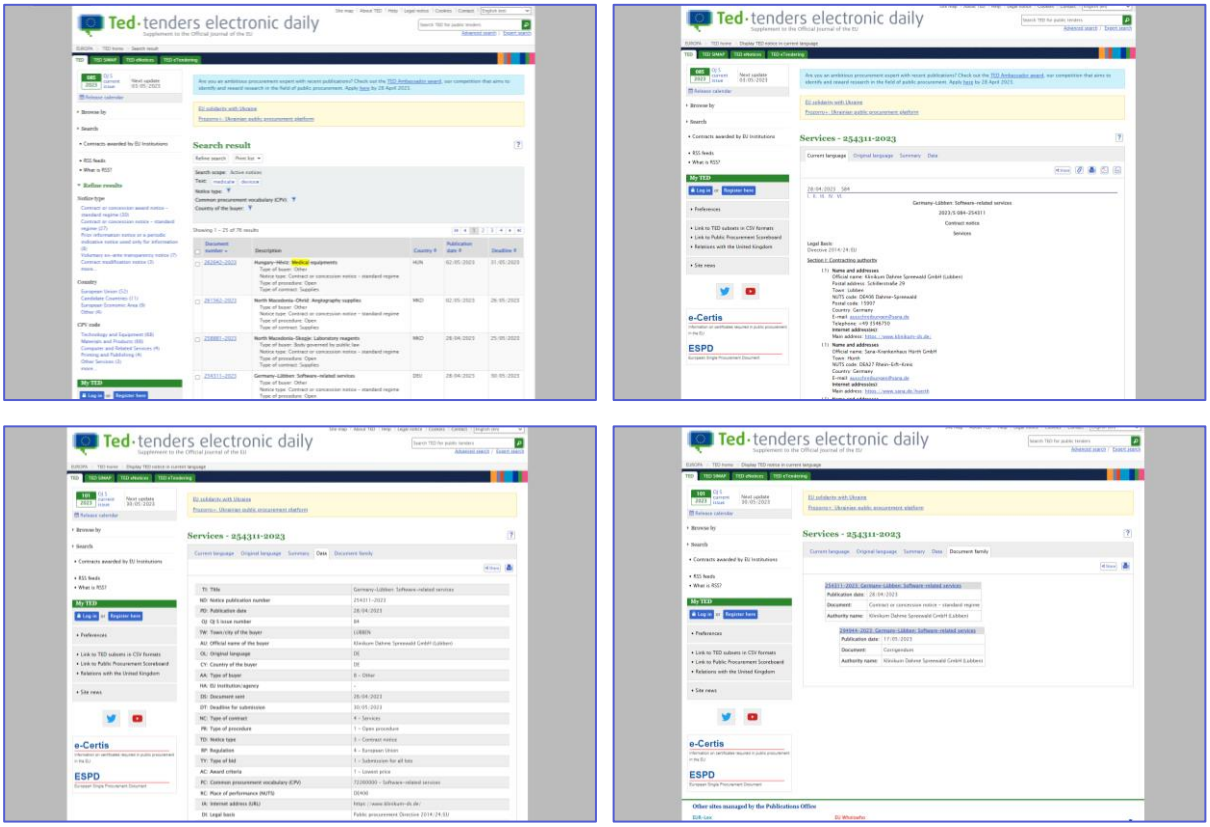



**Figure 35.** EU Tenders Example (TED)

# 5 Biomaterial Marketplace: Web applications, interfaces and digital advisors

The Biomaterial Marketplace aims to visually represent supplier organisations and their solutions, like products, services or documents. It operates in the areas of B2B (Business to Business) and B2G (Business to Government) – the main objective is to provide information about the suppliers and their biomaterial solutions and ease the contact with those suppliers. The Biomaterial Marketplace is realised as a dynamic web application. In the course of the holistic implementation, only current state-of-the-art technologies are used to guarantee a high-quality and modern platform as the end product. An analysis of technologies for various categories can be found in this chapter. Among the solution-oriented modules (like suppliers, products and services) the Biomaterial Marketplace will also come with digital advisors to assist users to identify their need and the best fitting solutions. The advisor results are based on several questions the user has to answer.

### User Interface

For the visual representation of biomaterial suppliers and their various solutions, the theme Limitless (https://demo.interface.club/limitless) is used as basis template. Limitless is a design template built with HTML, CSS, and JavaScript, designed to support developers for creating clean and responsive web applications. The template comes with a range of pre-designed pages and fully customisable components, including dashboards, forms, tables and charts. Limitless also includes features such as dark and light modes, multiple colour schemes and a modular architecture that makes it easy to add or remove features as needed. It also offers integration with popular UI frameworks like Bootstrap.

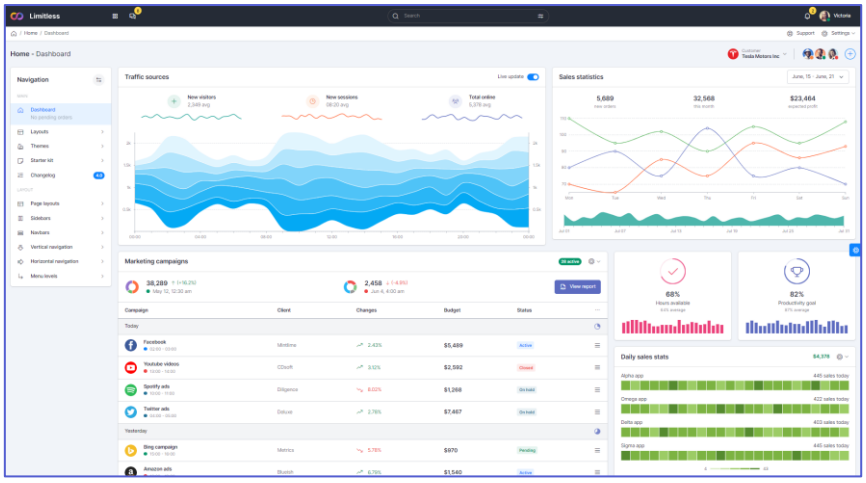



**Figure 36.** Limitless Theme

The Biomaterial Marketplace provides the same design approach for all entities (suppliers, products, services, etc.). Accordingly, for all entities, there are initially overview pages on which all existing entries are presented to the user. For all overview pages there are at least two display options - a list view and a grid view. There is also the possibility to filter and search the results according to different criteria. Each entity also has a detail page where additional information is presented to the user as well as related entities. For example, on the detail page of a supplier the user can also see the listed products and services of that supplier.

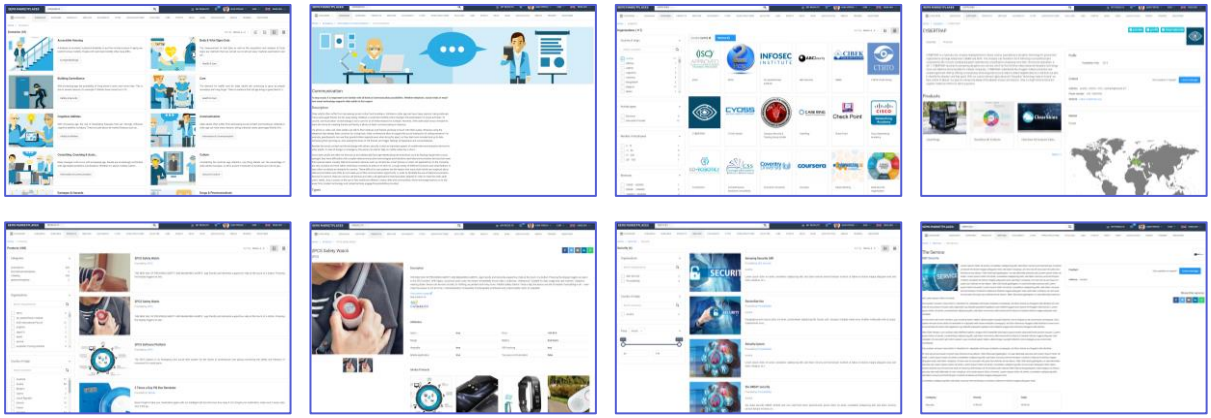


**Figure 37.** Demo User Interface

## Dynamic forms and filters

An essential feature for the user experience of an online platform are dynamic forms and filters. In this context, 'dynamic' means that the elements of the user interface are responding to the action of the user beyond the normal behaviour of web components (e.g., clicking a checkbox checks/unchecks the box). For example, if a user selects a checkbox of a filter option, the interface is immediately responding by updating the further filter options and corresponding counters for the results.

The example below (Figure 38Figure 38. Dynamic Forms and Filters) shows the scenario in which the user is filtering suppliers by selecting 'Austria' as country of origin, whereafter the further filter options for activity types are updated immediately.

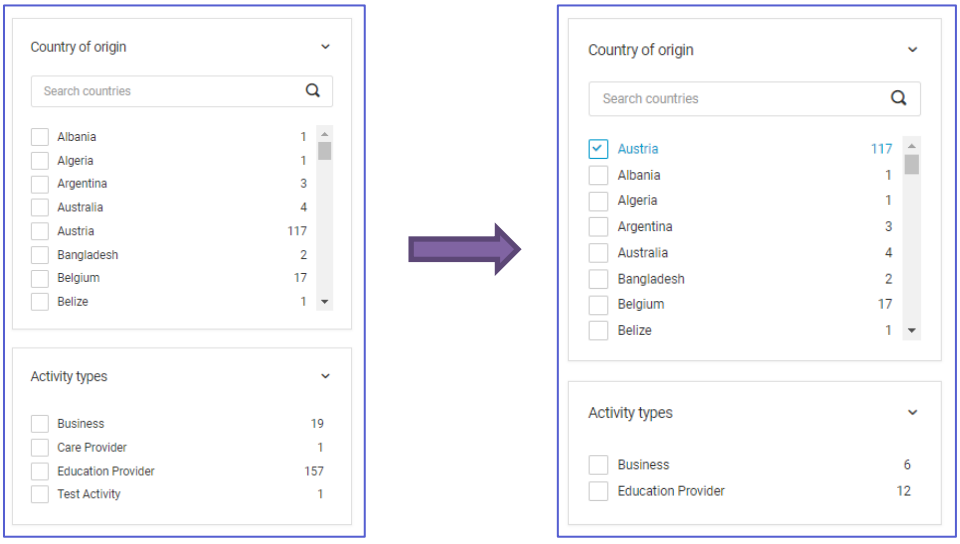



**Figure 38.** Dynamic Forms and Filters

## Digital Advisors

Digital advisors support users to find the best fitting solutions for their situation and needs. The advisor system combines a clean and intuitive user interface with intelligent queries and ranking methods that are used to find the best solutions for the user. The advisors are guiding the user step-by-step through the process by asking simple questions one after another about their situation and needs. During the process, the user has to answer multiple questions for example regarding price range or combability of biomaterials as well as simple yes or no questions. As soon as a defined set of questions has been

answered, the digital advisor redirects the user to a result page of solutions that suit best for the given needs.

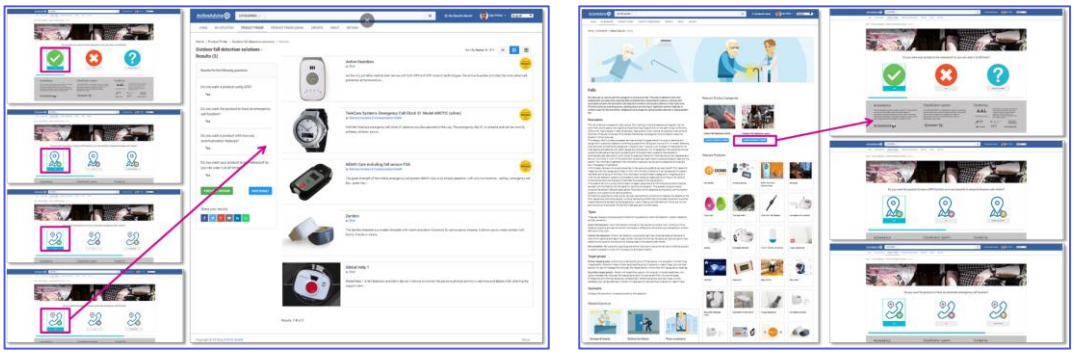


**Figure 39.** Advisor Workflow

### Model View Controller

To ease the modularly implementation of the Biomaterial Marketplace, the system is developed using the Model-View-Controller (MVC) pattern. This architectural pattern divides an application into three interconnected components: the model, the view and the controller. The model is responsible for the data structure within an application and is typically representing a table in the database, while the view presents information to the end user. Lastly, the controller is handling user inputs and updating both model and view accordingly.

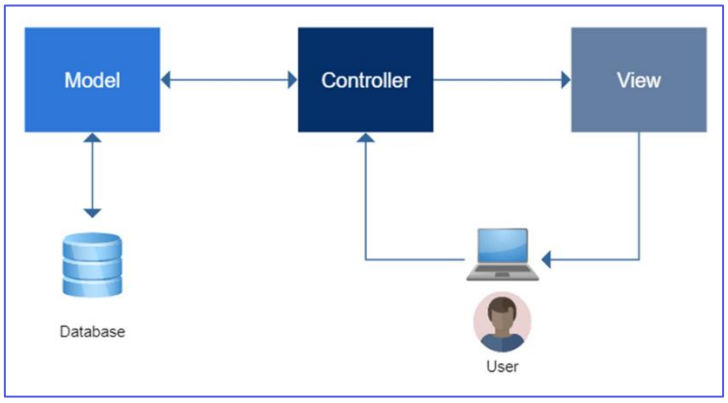



**Figure 40.** MVC Pattern

The separation of those components supports developers to read, maintain, scale and test code more efficiently. The MVC Pattern is commonly used via supportive framework. In the case of the Biomaterial Marketplace, the Laravel web framework is used - an open-source PHP web application framework.

### Intelligent Search

Intelligent search is a feature that enables highly accurate and fast search experience for users of the Biomaterial Marketplace. By implementing an intelligent search, it is possible to quickly search for biomaterials, suppliers and other information using natural language search queries.

The search engine uses advanced algorithms to analyse the search query and provide highly relevant search results based on factors such as search term frequency, proximity and context. The feature also includes autocomplete and suggestions such as the user types, enhancing the search experience and

helping users find what they need more quickly. Overall, the intelligent search feature aims to improve the usability and effectiveness of the Biomaterial Marketplace.

#### Analysis of technologies

As already noted, this chapter also includes an analysis of state-of-the art technologies used for online marketplace and platform solutions. The analysis of the technologies is divided into three parts: The first part deals with relevant system components (5.1), while the second part includes frontend-specific technologies (5.2). Additionally, an analysis for analytic components has been carried out (5.3). Regarding system components, the marketplaces were analysed in terms of Analytics, Content Delivery Network, Search Engine, Payment and Security. Relevant aspects for the frontend are User Interface and Visualisations, Maps, PDF Generators and Flip Book Brochures.

## 5.1 System Components

System components are tools and technologies that are essential to the operational capabilities of the marketplace system. The majority of these components remain invisible to the end user, residing in the underlying infrastructure and architecture of the system.

### 5.1.1 Analytics

Web analytics is the process of measuring, collecting, analysing and reporting website data to understand website performance and user behaviour. Analytic tools are able to track various metrics like the number of visitors, page views, bounce rate, session duration and conversion rate.

#### Google Analytics

Google Analytics is a web analytics service provided by Google. With its tracked metrics, it helps website owners to improve website content, design and user experience. As Google Analytics is a web-based service hosted by Google, it does not require advanced technical knowledge for an installation.

#### Matomo Analytics

Matomo Analytics is an open-source web analytics tool. One of the unique features of Matomo is – in addition to a cloud solution – the option of a self-hosted version. This requires additional IT infrastructure to install the self-hosted version on, but also provides users with complete ownership and control over their data – unlike other analytic tools that may share data with third-party providers.

### 5.1.2 Content Delivery Network

A content delivery network (CDN) is a network of geographically distributed servers. It ensures to keep response times (latency) as low as possible by using the server closest to the end user's locations. The distributed architecture also delivers a higher availability of the content.

#### Cloudflare

Cloudflare is a content delivery network tool that provides various performance and security-related features to websites. It offers security features like DDoS protection, web application firewall, SSL/TLS encryption and bot mitigation to safeguard websites against modern cyber threats.

#### Amazon CloudFront

Amazon CloudFront is a content delivery network tool offered by Amazon Web Services (AWS). It helps businesses to distribute their website content, videos, applications and APIs securely and quickly to users around the world. It also provides various security features like SSL/TLS encryption, field-level encryption and access controls to protect websites from various security threats.

#### Google Cloud CDN

Google Cloud CDN is a content delivery network (CDN) tool offered by Google. Cloud CDN is fully integrated with services like Compute Engine, Load Balancing and Cloud Storage to create a scalable and secure infrastructure for websites.

### 5.1.3 Search System

Search engines are software tools that allow users to search products, services or other entities on a marketplace platform. These search engines are designed to find relevant information quickly and efficiently. They use various algorithms and techniques to rank search results based on relevance, popularity, and other factors. Search engines are essential to the user experience of a marketplace and are critical to driving engagement and conversion rates.

#### Algolia

Algolia is a search-as-a-service platform that provides a fast search experience on their websites and mobile applications. It helps businesses to improve user engagement and conversion rates by delivering real-time search results that are tailored to each user's query and behaviour. Algolia uses a combination of machine learning and natural language processing to understand user queries and return the most relevant search results. It supports various search features like full-text search, faceted search, autocomplete and typo tolerance, making it suitable for various use cases like e-commerce, job boards and SaaS applications. Algolia offers different pricing plans, including a free version with limited search requests per month.

#### ElasticSearch

Elasticsearch is a distributed, open-source search and analytics engine that helps to store, search and analyse large volumes of data in real-time. It uses a document-based approach to store data in JSON format and provides various search and analysis features to quickly retrieve and analyse the data. Elasticsearch is highly scalable and can handle large volumes of data and complex search queries. ElasticSearch is a self-hosted search engine. This leads to a more complex setup as additional infrastructure and technical knowledge are required but also gives the opportunity to adapt it to specific needs – which is more difficult with cloud services.

#### Google Search

Google Search is a web search engine developed by Google that allows users to find information on the internet by searching web pages based on keywords and phrases. Google Search uses an algorithm called PageRank, which analyses the links between web pages to determine the relevance and authority of the web page. The algorithm takes various factors into account, including the relevance, popularity and authority of the web page, to determine the ranking of the search results. Google Search provides various search features like full-text search, image search, video search and news

search, making it suitable for various use cases. Additionally, Google Search provides various search tools like advanced search, filters and search operators to help users refine their search results.

### 5.1.4 Payment System

Payment services on marketplaces are software tools that allow buyers and sellers to exchange money for goods or services listed on a marketplace platform. These payment services typically provide a secure and efficient way to process payments, ensuring that buyers receive their purchases and sellers receive their payments in a timely manner. Payment services on marketplaces are typically integrated into the platform's user interface, making it easy for buyers and sellers to complete transactions without leaving the platform. They are critical to the success of a marketplace, as they enable the platform to monetise transactions.

#### Stripe

Stripe is a technology company that provides businesses with a set of APIs and tools to enable online payment processing and management. Stripe helps to accept payments from customers around the world, including credit cards, debit cards and various digital payment methods like Apple Pay and Google Pay. Stripe provides various features like subscription management, fraud detection and billing automation to manage payments and reduce operational costs.

#### Adyen

Adyen is a global payment platform that enables to accept payments from customers across different channels and geographies with a single solution. With Adyen, businesses are able to handle payments from various payment methods and digital wallets. Adyen also comes with additional features like risk management, fraud detection and reporting tools.

#### Invoice Ninja

Invoice Ninja is an open-source invoicing and billing software that allows businesses to create and send professional invoices, track expenses and accept payments online. Invoice Ninja helps businesses to streamline their invoicing and billing process, reduce administrative tasks and get paid faster. Invoice Ninja comes with functionalities like time tracking, project management, recurring invoicing and customizable invoice templates. It also provides various analytics and reporting tools to help businesses monitor their financial performance and optimize their revenue.

### 5.1.5 Security

Security tools are designed to protect web platforms and its users from various security threats and vulnerabilities. These tools typically provide features like user authentication, access control, data encryption and threat detection to ensure that the platform and its users are protected against malicious actors, data breaches and other security risks.

#### reCAPTCHA

reCAPTCHA is a security measure developed by Google that helps websites to prevent spam and automated bots from accessing their content or submitting forms by determining whether a user is a human or a bot. The most common type of reCAPTCHA is the "I'm not a robot" checkbox, which asks users to verify that they are not a bot by selecting the checkbox. In some cases, reCAPTCHA may ask

users to solve a challenge like identifying objects in an image or listening to an audio clip and typing out what they hear.

#### HSTS

HSTS [17] stands for HTTP Strict Transport Security. It is a security feature that instructs web browsers to only communicate with a website using HTTPS (HTTP Secure) connections. HSTS prevents security threats like downgrading attacks, where an attacker tries to force a website to use an unencrypted HTTP connection instead of HTTPS. It also helps to prevent man-in-the-middle attacks, where an attacker intercepts and modifies data exchanged between a user and a website. By enforcing the use of a secure connection, HSTS ensures the confidentiality and integrity of data transmitted over the web.

## 5.2 Frontend Components

Frontend components encompass a group of tools and libraries that serve to improve the visual representation and user experience of the marketplace frontend.

### 5.2.1 User Interface and Visualisation

User Interface and visualisation elements are instrumental in presenting information, facilitating easy navigation, and providing visual representations for better comprehension. By focusing on these aspects, the application can effectively serve the needs of users and promote the successful exchange and utilisation of biomaterials.

#### Bootstrap

Bootstrap is an open-source front-end framework that allows developers to create responsive and mobile-first websites easily. Bootstrap provides HTML, CSS and JavaScript components and templates that can be used to create modern and intuitive user interfaces. Bootstrap provides components like grid system, typography, forms, buttons, navigation and responsive utilities to help developers create consistent and scalable designs. Additionally, Bootstrap provides various plugins and extensions to enhance the functionality of the framework, making it suitable for various use cases. Bootstrap is widely used by developers and designers to create responsive and mobile-first websites and web applications and it is supported by a large and active community of developers who contribute to its development and improvement.

#### Data-Driven Documents (D3)

D3.js [16] is a widely used open-source JavaScript library for creating web-based data visualisations. It provides a flexible API to transform data into HTML, SVG, and CSS, making it a robust data visualisation framework. With D3.js, developers can create customizable visualisations and graphics from scratch, including complex multi-variate visualisations that are fully connected and brush-able. It is the most popular library for data-driven documents and can implement everything from simple bar-charts to advanced visualisations.

#### Chart.js

Chart.js is a simple yet very flexible JavaScript library for data visualisation that is popular with web designers and developers. It is a good basic solution for those who do not need many chart types and customisation features, but want their charts to look neat, clear and informative at a glance.

### 5.2.2 Maps

The integration of interactive maps into an online marketplace enhances the user experience and provides valuable geospatial information. It allows users to visually explore product or service locations, and navigate to desired destinations. By integrating web mapping services, online marketplaces can offer users a spatial perspective, enabling them to make informed decisions based on geographic context.

#### Google Maps

Google Maps is a web mapping service developed by Google that provides users with detailed maps and satellite imagery of locations around the world. It allows users to search for specific addresses, businesses and points of interest. The service also provides real-time traffic updates, directions and transit information to help users navigate their way to their destination. Additionally, Google Maps provides various APIs that allow developers to integrate mapping and location-based services into their websites and applications.

#### Leaflet

Leaflet is an open-source JavaScript library that allows developers to create interactive maps and geospatial applications on the web. Leaflet offers features like zooming, panning, tile layers, markers and popups to help developers create dynamic and customisable maps. Leaflet is highly modular and can be easily extended with various plugins and extensions to add more functionality to the maps. Leaflet is lightweight, fast and mobile-friendly, making it suitable for various use cases.

#### OpenStreetMap

OpenStreetMap (OSM) is a collaborative project that provides free and editable maps of the world, similar to commercial map providers like Google Maps. OSM is built and maintained by a community of volunteers who contribute data and maintain the maps. The data is collected using tools like GPS devices, aerial imagery and government datasets. OSM provides features like detailed street maps, satellite imagery, routing and navigation. It also supports multiple map styles that can be customised for applications. Additionally, OSM provides various APIs that allow developers to access and use the data in their own applications, enabling them to create custom mapping solutions. OSM is widely used by individuals, businesses, and organisations for a wide range of applications, including navigation, logistics, emergency response, and urban planning. OSM is free to use and distribute under an open license, making it a valuable resource for the global community.

### 5.2.3 PDF Generators

PDF generators allow users to generate and download PDF documents based on marketplace-related content. This feature enables users to create printable versions of product listings, invoices, receipts, or other relevant information in a standardised and easily shareable format.

#### PrinceXML

PrinceXML is a commercial software tool that allows developers and designers to convert HTML and CSS documents into high-quality PDFs, including complex documents with multiple columns, tables and graphics. PrinceXML supports various CSS3 features, including flexible box layout, grid layout and multicolumn layout. The tool also provides advanced typography features like hyphenation, ligatures, and kerning. Additionally, PrinceXML comes with features like headers, footers, page numbering and table of contents to facilitate the creation of professional-looking documents.

#### DOMPDF

DOMPDF is an open-source library for generating PDF documents from HTML and CSS. It is written in PHP and can be used to create PDF files from various data sources, including web pages, databases and other document formats. DOMPDF uses the PHP-based HTML/CSS rendering engine to convert the HTML and CSS documents into PDF format, which can then be saved or displayed in the web browser. DOMPDF supports various features like inline PHP support, CSS 3 support, font embedding, image support and PDF encryption. It also provides a simple API for developers to generate PDF documents programmatically, making it easy to integrate into existing PHP projects.

#### wkhtmltopdf

wkhtmltopdf is a command-line tool that converts HTML and CSS documents into PDF files. It is based on the WebKit rendering and can be used to generate high-quality PDF documents from various data sources, including web pages, local HTML files and other document formats. wkhtmltopdf supports various features like page size and orientation, header and footer customisation, table of contents generation and JavaScript execution. Furthermore, wkhtmltopdf is an open-source solution.

### 5.2.4 Flip Book Brochures

Flip book brochures refer to interactive and visually appealing digital documents that emulate the experience of flipping through a physical brochure. These brochures are designed to engage users with immersive and interactive content. In the context of online marketplaces, digital flipbook brochures serve as dynamic marketing tools to showcase products or services in a visually compelling way.

#### Real3D Flipbook jQuery Plugin

Real3D Flipbook is a commercial jQuery plugin by creativeinteractivemedia that allows users to create realistic and interactive flipbooks in HTML5 format. It provides a smooth and realistic page-flipping effect that mimics the look and feel of a physical book, and supports various customisation options like book size, page thickness, cover styles and zooming options. Real3D Flipbook can be used to display images, PDFs and other documents in a flipbook format. The plugin is fully responsive and works on all modern devices and browsers, including mobile devices. It also provides various controls and tools for navigating and interacting with the flipbook, such as bookmarks, thumbnails, search and sharing options.

#### 3D Flipbook JavaScript Plugin

3D Flipbook is a commercial JavaScript plugin for creating 3D flipbooks with realistic page-flipping animations. It can display PDFs, images and other documents in an interactive flipbook format that can be viewed on desktop and mobile devices. The plugin uses WebGL technology to create a 3D

environment that simulates the look and feel of a physical book, with customisable features. 3D Flipbook also supports various customisation options for page flip animations, shadows, reflections and lighting effects. The plugin provides a user-friendly interface for navigating and interacting with the flipbook, including controls for zooming, panning, and rotating the book.

## 5.3 Analytics Components

The Biomaterial Marketplace utilises various recommendation systems/approaches for engaging with the users of the marketplace and to tailor information based on their context and the needs of the user. This information is largely based on a remix of existing open data provided by the EU, namely:

- European **funding** calls and topics – In order to identify funding opportunities under relevant funding schemes and topics/domains.
- European **organisations** – In order to foster collaboration and/or to identify the competitive landscape of the users.
- European **research project** – In order to facilitate engagement across research activities.
- European **tenders** – In order to identify potential buyers and identify business opportunities.

The following section provide an overview of the potential of these data sources, and how a mapping between user needs and opportunities could be realised.

### 5.3.1 Assumptions, Availability of Information and Limitations

The marketplace generally suffers from a cold start problem, meaning that, at least initially, it has only a limited amount of information about the individual users, with which it may operate and provide recommendations on. This level of information depends on the stage of the involvement of the user, and the consortium assumes that there are the following levels of information that the users could have provided:

- The initial stage consists of providing at least their location (primarily country is of relevance), their organisation, and in which domain this organisation is operating in. This at least allows for the system to limit the scope geographically (if this limitation applies) and to classify the users within a specific domain, and may be used to match them, if an appropriate mapping already exists or can be accurately created (e.g., mapping locality to Nomenclature of Territorial Units for Statistics (NUTS)).
- In case the user is a supplier, they should already have products registered with various attributes and classifications, which may be used to further narrow down the recommendation to e.g., specific products (in direct matches in the recommended entities).
- On a longer term, the user's previous preferences may be noted, and used to identify future upcoming opportunities, based on their similarity.

Further, the information provided by the user should be matched with the various datasets, their content and their meta-data.

#### Limitation of the Considered Datasets

The considered datasets generally provide a measurable volume of data, in terms of entities, as well as metadata for each of the entities. However, the information that is provided is largely orthogonal to each other as well as to what is known about the user, and what is reasonable to expect the user to

provide. For example, an organisation's project participation count is typically quite low, the engagement of an organisation in European research topics and calls is unknown and the general background of the user's organisation may only be known to a limited degree. Therefore, alternative approaches need to be considered for providing recommendations to the users, and ensuring that the content they are presented with is relevant for them and their activities. Further, the marketplace does, in most cases, neither have access to user history/preference, or the available information provides limited context, as the consortium is looking at providing relevant results based on future events, which reduces the applicability of methods such as link predictions. Table 26 provides an overview of the datasets and which attributes, "out of the box", are provided that could be used as indicators of relevance, and the topics is further elaborated in the following subsection.

**Table 26**. Available annotations for the considered datasets

| Domain | Classification | Description |
|---|---|---|
| **Research topics** | Built in labels and classification (keywords, flags, tags) | European funding topics are, in addition to a general description and meta-data, described/annotated with keywords, tags and flags, providing an indicator of the domain and expected focus of the topic. This may be exploited to filter out relevant topics, e.g., based on various related topics. |
| **Research projects** | EuroSciVoc, keywords | Research projects are annotated directly with the EuroSciVoc, hence may be directly segmented based on their "Field of science", as well as additional information may be derived from the original classification of the topics under which the project was funded under. |
| **Organisations in research** | Indirect EuroSciVoc | Organisations that have participated in European research project are labelled by a) their own classifications, that they themselves may have provided, and by the tags derived from the projects in which they have participated in, hence providing, at least to some extent, an indicator for their area of expertise and/or area of operation. |
| **Public tenders** | Common Procurement Vocabulary (CPV), NUTS | The CPV provides a broad classification method that ranges all the way from high level domains, e.g., "Health and social work services" being a top-level classification, down to individual products such as insulin or prosthesis. |

### Funding & Tender Opportunities

Funding and tenders, mainly represented by HORIZON calls and topics, provide the users with opportunities for identifying funding for further development of their ideas and product, potentially improving their applicability and quality. Generally, funding topics are structured vertically and horizontally, according to topical clusters, such as health, climate, and domains/mission, such as cancer as well as various climate-oriented missions. This naturally provides a broad range of filters for narrowing down specific topics; however, the options provided may not necessarily resonate with the

terminology of the users, thus, may need to be further refined. Besides the natural relationships of topics, the topics descriptions could also provide specific annotations that may be beneficial, namely, tags, keyword and flags, though they are not used very consistently, may still provide further context.

#### European Research Projects and Research Results

With a budget of nearly 80 billion Euro, H2020 alone funded +35.000 projects. As projects are funded under specific topics, as mentioned above, they may somewhat derive their topical properties and classifications from them directly. At least initially, only a short description of the project is available for each project, however, as the project progresses, CORDIS is adding additional project results, such as deliverables, publications and project updates, which may be further used to narrow down the classification of the project and in which domains it operates in. When consolidated, this information may be used to provide the users with recommendations on what is a) ongoing within their domain, and b) to explore uptake opportunities.

#### European Research Organisations

As an extension of European research projects, the project meta-data also contains information about the organisations that are involved in a given project. While the role of the individual organisations is not directly specifying, making the contribution of the specific organisation somewhat blurred, it provides indicators for which domains and areas an organisation may be involved in, at least when it comes to their research-related activities. Implicitly, all the information provided about a given project may be projected onto the participating organisation, including the inaccuracies this may introduce.

#### Tenders Electronic Daily

Tenders Electronic Daily (TED) provides daily opportunities, around 3.000 daily postings, for connecting suppliers with potential buyers, Europe wide. The postings consist of a collection of public, national and EU level, tenders across whole of EU, provided in a standardized format. While some of the primary foci are to provide transparency and support investigative journalism, at the same time, it serves as a comprehensive collection of upcoming and current opportunities for goods and services that the user's organisation may be providing. The offered tenders, however, are somewhat limited in what meta data they provide, at least in terms of their domain and categorical content, but, as mentioned, are annotated with at least the Common Procurement Vocabulary (CPV)[7], and a short description in the original language, or highly variable quality. Additional information is typically provided by the original procurer, such as various specification document, however, this may not always be easily accessible when targeting automatic processing of the data on a large scale.

### 5.3.2 Matching and Recommendations

The project will evaluate various approaches for how to facilitate the matching between the needs of the users and the available opportunities, each at an increasing level of complexity. Inherently, the considered datasets already provide a certain level of labelling and classifications of the provided entities, e.g., following the EuroSciVoc taxonomy and the Common Procurement Vocabulary (CPV) for classification of European research activities and public procurement opportunities, respectively.

[7] https://single-market-economy.ec.europa.eu/single-market/public-procurement/digital-procurement/common-procurement-vocabulary_en

While these approaches have their limitations, e.g., primarily due to mislabelling and/or less than accurate labelling, they provide a basis that may be improved upon, which will be evaluated.

For providing advisory to the users, and to provide them with relevant/recommended results, the capabilities of matching between the user needs and the available resources are necessary. Hence, the consortium is interested in matching and ranking entities. As pointed out, most of the considered entities are not directly suited for conventional prediction, based on historical data, due to their event-driven nature, so the approach needs to be based on more generalizable approaches where user needs and feature provided by the entities may be matched.

### General Matching

As the two mentioned taxonomies, EuroSciVoc and CPV, are already well defined, they may be used directly, as in provided by the user, to look up entities that are labelled with them, thus, providing direct matches, if they exist, based on these attributes. As a further support, the marketplace may recommend similar entities, based on the overlap of their mutual annotations. While matching between labels may provide matches in cases where direct matches actually exist, the approach is limited by its binary nature; e.g., given a combination of keywords, results can only be found if an exact match exists.

### Extended Matching, Ranking and Classification

As basic matching has some limitations due to its rigidness, alternative approaches will be explored. Generally, this means that requirements for the alternative approaches need to be support at least some levels of fuzziness, where matches are identified based on a degree of similarity instead, and ranked against each other, as well as broadening out the scope of the annotations of the individual entities by extracting implicit relations.

#### Full-Text Search and Embeddings

Out of the box, Elasticsearch provides full-text search capabilities, that by itself enables keyword-based search and ranking (e.g., TF-IDF), thus allowing for a basic search approach, beyond pure attribute matching. Furthermore, newer revisions of Elasticsearch also provide support for indexing of dense vectors, and support for, e.g., kNN based queries. Combined with an appropriate vectorization approach, this may further improve the quality of the presented results, by extracting the relevant features from the collected entities as well as any relations that may be available as this approach is capable of being less restrictive than full-text search alone.

#### Keyword Extraction and Natural Language Processing

While full-text search provides a “low-effort“ entry to provide search results to the user, it is often not capable of taking the full context of the text into account, which may result in providing results that are somewhat relevant, but also may contain noisy outliers. Rather, various natural language processing techniques could be applied that attempt to encapsulate the information of entities that provide a textual description. Here, models exist that focus on medical terminology, as also outlined in Section 3, and may be used to extract named entities (such as products, medicine, names of diseases) and also generic terms that attempt to be representative for the text itself. These may in turn be applied as filters or to position the entities in knowledge graphs as described below.

### Similarity Search Encompassing Textual, Categorical and Numerical Attributes

Out of the box, the datasets provide mostly categorical attributes which allow for the application of various set similarity measures. Tools such as SimSearch, developed by the SmartDataLake project, provide a flexible implementation of a range of similarity metrics and support multi-attributes combinations, including categorical, textual/keyword based, numerical, spatial and temporal attributes. The component may run as a stand-alone component, ingested with the appropriate data, or be used as an intermediate component on top of an existing solution, e.g., being connected directly to an existing Elasticsearch instance. However, additional approaches should be considered as well as optimisation of the provided, by the data, attributes.

### Graph Based Approaches

Most of the considered datasets are relational by nature; projects have project partners, publications (that have been published at venues), in extension, projects are funded under topics which in turn fall under overarching topics/calls focusing on various domains. When accurately modelling these relations as a graph, using, e.g., a graph database such as Neo4J, it provides additional entry points for finding a domain and/or topical similarities between the relevant information and the, most likely, limited information that is available about the user, by traversing the graph accordingly. This would also enable the marketplace to show similar entities based on where they are located in the graph, based on the naturally given properties (relationships). E.g., both EuroSciVoc and CPV are organized hierarchically, hence while an entity may not be directly relatable, levels of separation may be introduced to provide entities that fall into either a parent or neighbouring node. Here, one option is to annotate the collected entities based on the named entities that have been recognized in the description and the meta-data, and then identifying commonalities between the user and the entities though traversal of the graph.

**Table 27**. Potential candidates for expanding entity annotations

| Domain | Additional Candidates |
|---|---|
| **Research topics** | ▪ (Text) Topic classification<br>▪ Keyword extraction |
| **Research projects** | ▪ Topic classification from descriptions<br>▪ Keyword extraction<br>▪ Publication graphs |
| **Organisations in research** | ▪ Would need additional databases/source |
| **Public tenders** | ▪ (Text) Topic classification<br>▪ Medical named entity recognition w/o knowledge graphs |

## 5.4 Components Overview

The following two tables (Table 28 and Table 29) provide an overview of the system and front-end components and its categories as described in the previous chapters (5.1 and 5.2).

**Table 28.** Overview Table (System Components)

| | |
|---|---|
| **Analytics** | |
| Analytics | **Google Analytics**<br>Google Analytics is a web analytics service by Google. With its tracked metrics it helps website owners to improve website content, design and user experience. |
| matomo | **Matomo Analytics**<br>Matomo Analytics is an open-source web analytics tool. One of the unique features of Matomo is – in addition to a cloud solution – the option of a self-hosted version. |
| **Content Delivery Network** | |
| CLOUDFLARE | **Cloudfare**<br>Cloudflare is a content delivery network tool that provides various performance and security-related features to websites. |
| Amazon CloudFront | **Amazon CloudFront**<br>Amazon CloudFront is a content delivery network tool offered by Amazon Web Services (AWS). |
| Cloud CDN | **Google Cloud CDN**<br>Google Cloud CDN is a content delivery network (CDN) tool offered by Google. |
| **Search System** | |
| algolia | **Algolia**<br>Algolia is a search-as-a-service platform that provides a fast search experience on their websites and mobile applications. |
| elasticsearch | **ElasticSearch**<br>Elasticsearch is a distributed, open-source search and analytics engine that helps to store, search and analyse large volumes of data in real-time. |
| Google Search | **GoogleSearch**<br>Google Search is a web search engine developed by Google that allows users to find information on the internet by searching web pages based on keywords and phrases. |
| **Payment System** | |
| stripe | **Stripe**<br>Stripe is a technology company that provides businesses with a set of APIs and tools to enable online payment processing and management. |
| adyen | **Adyen**<br>Adyen is a global payment platform that enables to accept payments from customers across different channels and geographies with a single solution. |
| InvoiceNinja | **Invoice Ninja**<br>Invoice Ninja is an open-source invoicing and billing software that allows businesses to create and send professional invoices, track expenses and accept payments online. |
| **Security** | |
| reCAPTCHA | **reCAPTCHA**<br>reCAPTCHA is a security measure developed by Google that helps websites to prevent spam and automated bots from accessing their content or submitting forms |
| HSTS Secured With HSTS | **HSTS**<br>HSTS stands for HTTP Strict Transport Security. It is a security feature that instructs web browsers to only communicate with a website using HTTPS (HTTP Secure) connections. |

**Table 29.** Overview Table (Frontend Components)

| | |
|---|---|
| **User Interface and Visualisation** | |
|  | **Bootstrap**<br>Bootstrap is an open-source front-end framework that allows developers to create responsive and mobile-first websites easily. |
|  | **Data-Driven Documents (D3)**<br>D3.js is a widely used open-source JavaScript library for creating web-based data visualisations. |
|  | **Chart.js**<br>Chart.js is a simple yet very flexible JavaScript library for data visualisation that is popular with web designers and developers. |
| **Maps** | |
|  | **Google Maps**<br>Google Maps is a web mapping service developed by Google that provides users with detailed maps and satellite imagery of locations around the world. |
|  | **Leaflet**<br>Leaflet is an open-source JavaScript library that allows developers to create interactive maps and geospatial applications on the web. |
|  | **OpenStreetMap**<br>OpenStreetMap (OSM) is a collaborative project that provides free and editable maps of the world, similar to commercial map providers like Google Maps. |
| **PDF Generators** | |
|  | **PrinceXML**<br>PrinceXML is a commercial software tool that allows developers and designers to convert HTML and CSS documents into high-quality PDFs. |
|  | **DOMPDF**<br>DOMPDF is an open-source library for generating PDF documents from HTML and CSS. |
|  | **wkhtmltopdf**<br>wkhtmltopdf is a command-line tool that converts HTML and CSS documents into PDF files. |
| **Flip Book Brochures** | |
|  | **Real3D Flipbook jQuery Plugin**<br>Real3D Flipbook is a commercial jQuery plugin by creativeinteractivemedia that allows users to create realistic and interactive flipbooks in HTML5 format. |
|  | **3D Flipbook JavaScript Plugin**<br>3D Flipbook is a commercial JavaScript plugin for creating 3D flipbooks with realistic page-flipping animations. |

# 6 Conclusion

The aim of this deliverable was to lay the foundation for the different components by outlining the framework-selection process, and describing the frontend-components, the data-collection and the annotation process.

Hence, the deliverable described the two main technical parts: the (1) Biomaterial Database and the (2) Biomaterial Marketplace.

### Biomaterial Database

The report detailed the usage of advanced NLP techniques to enhance information retrieval, categorisation, and semantic indexing of biomaterial-related content. Data sources, data flow, content selection, and pre-processing were key considerations throughout the process. Techniques such as topic modelling, text similarity, text classification, clustering, named entity recognition (NER), and metadata/entity normalisation were employed, supported by iterative improvements through guidelines, annotations, and machine learning models. Annotation tools, software, and models were presented, along with the initial design of the database management system.

In addition, the report emphasised the importance of pre-processing as a fundamental step in the NLP pipeline. The specific steps for the Biomaterial Database included topic modelling, text similarity, text classification and clustering, NER, automatic acquisition of hyponyms, and metadata and entity normalisation. Furthermore, possible frameworks and supporting libraries for the development were highlighted. Lastly, technologies and techniques for interactive visualisation in the web-interface to the database were presented.

Overall, the report provides a comprehensive overview of the development process and status of the Biomaterial Database, showcasing the utilisation of advanced NLP techniques. It lays a solid foundation for the successful development and management of the Biomaterial Database.

### Biomaterial Marketplace

The elaborated technical approach for the Biomaterial Marketplace orients itself on the three identified meta use cases: administration panel, marketplace and digital advisor. An important part of the technical system are the used data sources. While the core modules (suppliers, products, services, scenarios, documents, news, events) are all entered manually via the administration panel, the extension modules (EU projects, EU project partners, EU funding, EU tenders) retrieve data from external sources, provided by the European Union and/or its institutions. Regardless of the source, all the data flows into the technical system afterwards, where it is processed, stored and prepared to be accessible through an API endpoint. The information presented on the application frontend is then directly received from these endpoints.

Among the technical approach for the marketplace system, the web application and its user experience were covered in this deliverable. In addition to some general aspects like the user interface or the used development pattern, some special features were outlined in the scope of this section as well. Moreover, a broad analysis of state-of-the art components and technologies for marketplace solutions was carried out for various categories. The categories were divided in three different sub-sections:

system components, frontend components and analytics components. The following categories were analysed for the mentioned sub-sections:

- System components:
  - Analytics
  - Content Delivery Network
  - Search System
  - Payment System
  - Security
- Frontend components:
  - User Interface and Visualisation
  - Maps
  - PDF Generators
  - Flip Book Brochures
- Analytics Components:
  - Matching and Recommendations

### Upcoming activities

The deliverable, encompassing the results, concepts, tools, and processes introduced, will play a pivotal role in the subsequent activities of WP3 (Phase 2). These activities include the design and specification of the backend-components for the Biomaterial Database and the Biomaterial Marketplace. Moreover, the outputs of this deliverable will provide valuable support for activities in WP4, primarily for D4.3, which involves describing the implementation of frontend interfaces using modern web application frameworks, and D4.4, which focuses on the integration of relevant biomaterial data and entities.

By serving as the foundation for the upcoming work in WP3 and WP4, this deliverable will enable the detailed specification of data structures and formats, together with as the continuous design of the systems. Furthermore, it will support the iterative development of the systems, ensuring that they meet the objectives of the project effectively.

# References


[1] Corvi, J., McKitrick, A., Fernández, J., Fuenteslópez, C., Gelpi, J., Ginebra, M.-P., Capella-Guitierrez, S., Hakimi, O. DEBBIE: the open access database of experimental scaffolds and biomaterials with an automated information retrieval pipeline. Manuscript in preparation (2022).

[2] Corvi, J., Fuenteslópez, C., Fernández, J., Gelpi, J., Ginebra, M.-P., Capella-Guitierrez, S., Hakimi, O. The biomaterials annotator: a systemfor ontology-based concept annotation of biomaterials text. In:Proceedings of the Second Workshop on Scholarly DocumentProcessing, pp. 36–48. Association for Computational Linguistics, Online (2021). https://www.aclweb.org/anthology/2021.sdp-1.5

[3] Hakimi, O., Gelpi, J., Krallinger, M., Curi, F., Repchevsky, D., Ginebra, M.-P. The devices, experimental scaffolds, and biomaterials ontology (deb): A tool for mapping, annotation, and analysis of biomaterials' data. Adv. Funct. Mater. (2020)

[4] Y. Gu *et al.*, "Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing," *ACM Trans. Comput. Healthcare*, vol. 3, no. 1, pp. 1–23, Jan. 2022, doi: 10.1145/3458754.

[5] E. Alsentzer *et al.*, "Publicly Available Clinical BERT Embeddings." arXiv, Jun. 20, 2019. doi: 10.48550/arXiv.1904.03323.

[6] J. Lee et al., "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, Feb. 2020, doi: 10.1093/bioinformatics/btz682.

[7] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A Pretrained Language Model for Scientific Text." arXiv, Sep. 10, 2019. doi: 10.48550/arXiv.1903.10676.

[8] M. Grootendorst, "keybert: KeyBERT performs keyword extraction with state-of-the-art transformer models." Accessed: May 23, 2023. Available: https://github.com/MaartenGr/keyBERT

[9] T. Schopf, S. Klimek, and F. Matthes, "PatternRank: Leveraging Pretrained Language Models and Part of Speech for Unsupervised Keyphrase Extraction," in *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2022, pp. 243–248. doi: 10.5220/0011546600003335.

[10] Explosion, "spacy: Industrial-strength Natural Language Processing (NLP) in Python." Accessed: May 23, 2023. Available: https://spacy.io

[11] A. I. for A. Intelligence, "scispacy: A full SpaCy pipeline and models for scientific/biomedical documents." Accessed: May 23, 2023. [Online]. Available: https://allenai.github.io/SciSpaCy/

[12] P. Team, "torch: Tensors and Dynamic neural networks in Python with strong GPU acceleration." Accessed: May 23, 2023. [Online]. Available: https://pytorch.org/

[13] The Hugging Face team (past and future) with the help of all our contributors (https://github.com/huggingface/transformers/graphs/contributors), "transformers: State-of-the-art

Machine Learning for JAX, PyTorch and TensorFlow." Accessed: May 23, 2023. Available: https://github.com/huggingface/transformers

[14] G. Inc, "tensorflow: TensorFlow is an open source machine learning framework for everyone." Accessed: May 23, 2023. [Online]. Available: https://www.tensorflow.org/

[15] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J. Tsujii, "brat: a Web-based Tool for NLP-Assisted Text Annotation," in *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, France: Association for Computational Linguistics, Apr. 2012, pp. 102–107. Accessed: May 24, 2023. [Online]. Available: https://aclanthology.org/E12-2021

[16] M. Bostock, V. Ogievetsky, and J. Heer, "$D^3$ data-driven documents," IEEE transactions on visualization and computer graphics, vol. 17, no. 12, pp. 2301–2309, 2011.

[17] J. Hodges, C. Jackson, and A. Barth, HTTP Strict Transport Security (HSTS), Request for Comments. RFC Editor, doi: 10.17487/RFC6797. [Online]. Available: https://www.rfc-editor.org/info/rfc6797

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, May 24, 2019. doi: 10.48550/arXiv.1810.04805.

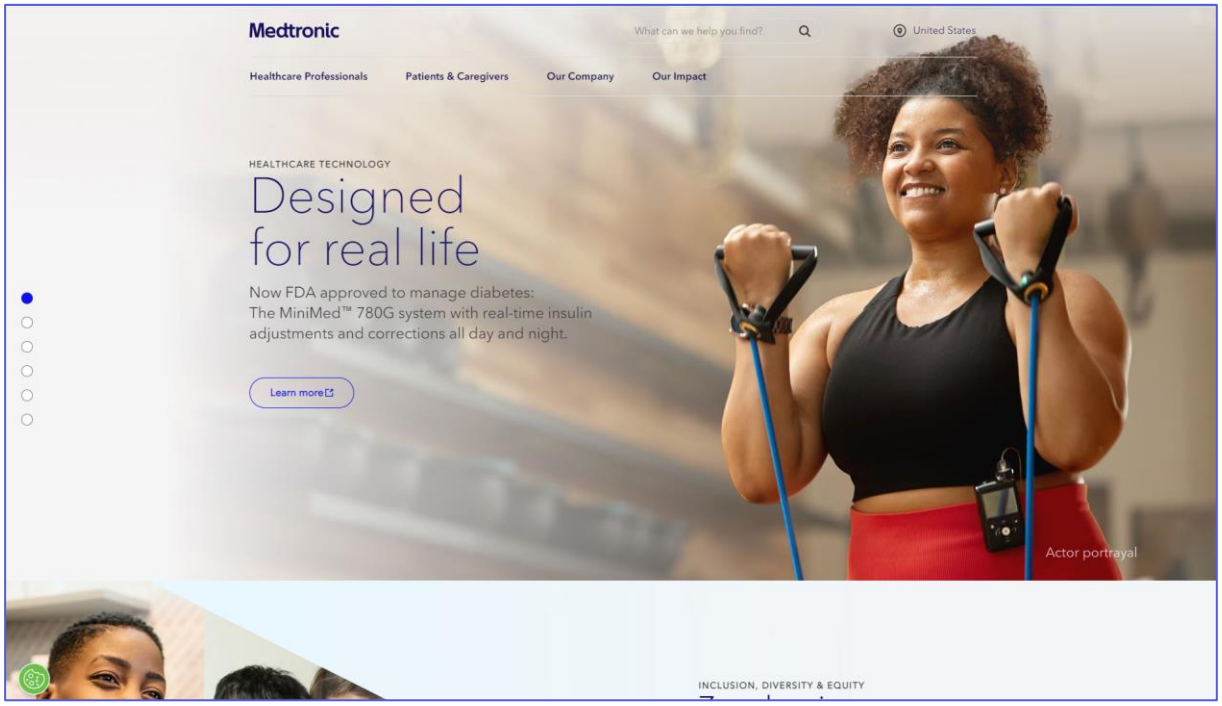
[19] "Prodigy: A new tool for radically efficient machine teaching · Explosion," *Explosion*, Aug. 04, 2017. https://explosion.ai/blog/prodigy-annotation-tool-active-learning (accessed May 30, 2023).

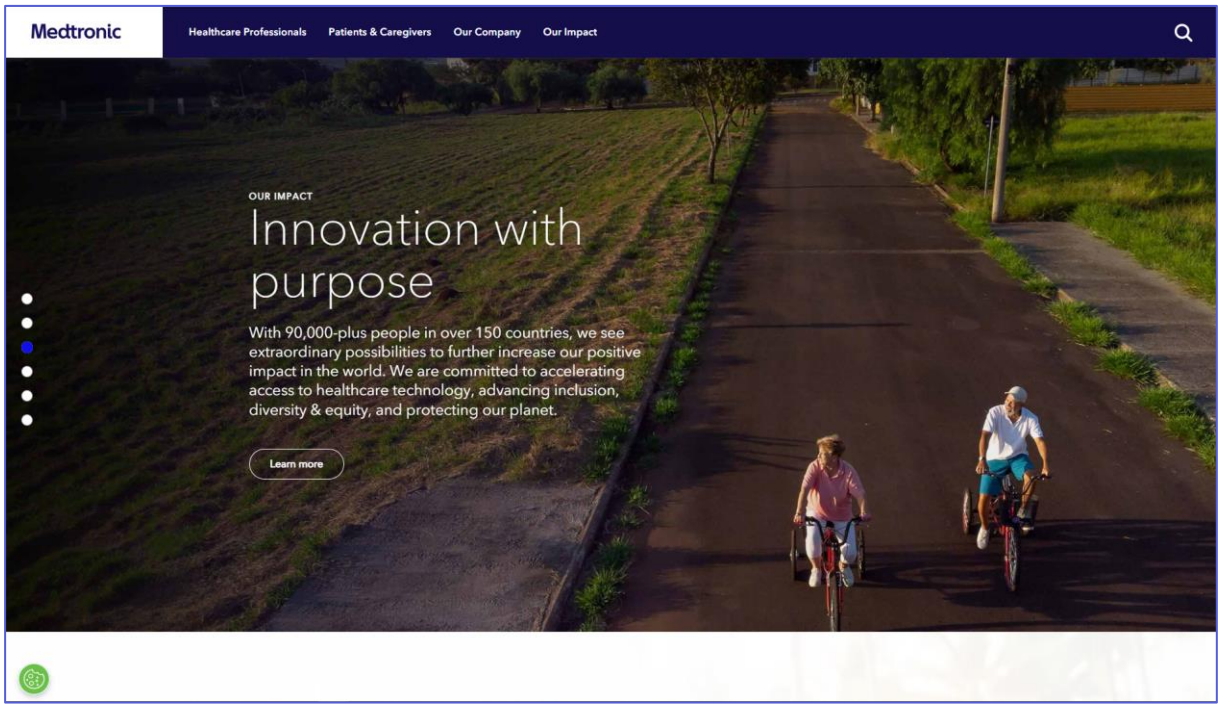
## Web References

| | |
|---|---|
| D3.js | d3js.org |
| Chart.js | chartjs.org |
| Bootstrap | getbootstrap.com |
| jQuery | jquery.com |
| Google Analytics | analytics.google.com |
| Matomo | matomo.org |
| Cloudflare | cloudflare.com |
| Amazon CloudFront | aws.amazon.com/cloudfront |
| Google Cloud CDN | cloud.google.com/cdn |
| Algolia | algolia.com |
| ElasticSearch | elastic.co/elasticsearch |
| Google Search | google.com/webhp |
| Stripe | stripe.com |
| Adyen | adyen.com |
| Invoice Ninja | invoiceninja.com |
| reCAPTCHA | google.com/recaptcha |
| Google Maps | google.com/maps |
| Leaflet | leafletjs.com |
| OpenStreetMap | openstreetmap.org |
| PrinceXML | princexml.com |
| DOMPDF | github.com/dompdf |
| wkhtmltopdf | wkhtmltopdf.org |
| Real3D Flipbook jQuery Plugin | real3dflipbook.com |
| 3D Flipbook JavaScript Plugin | 3dflipbook.net |

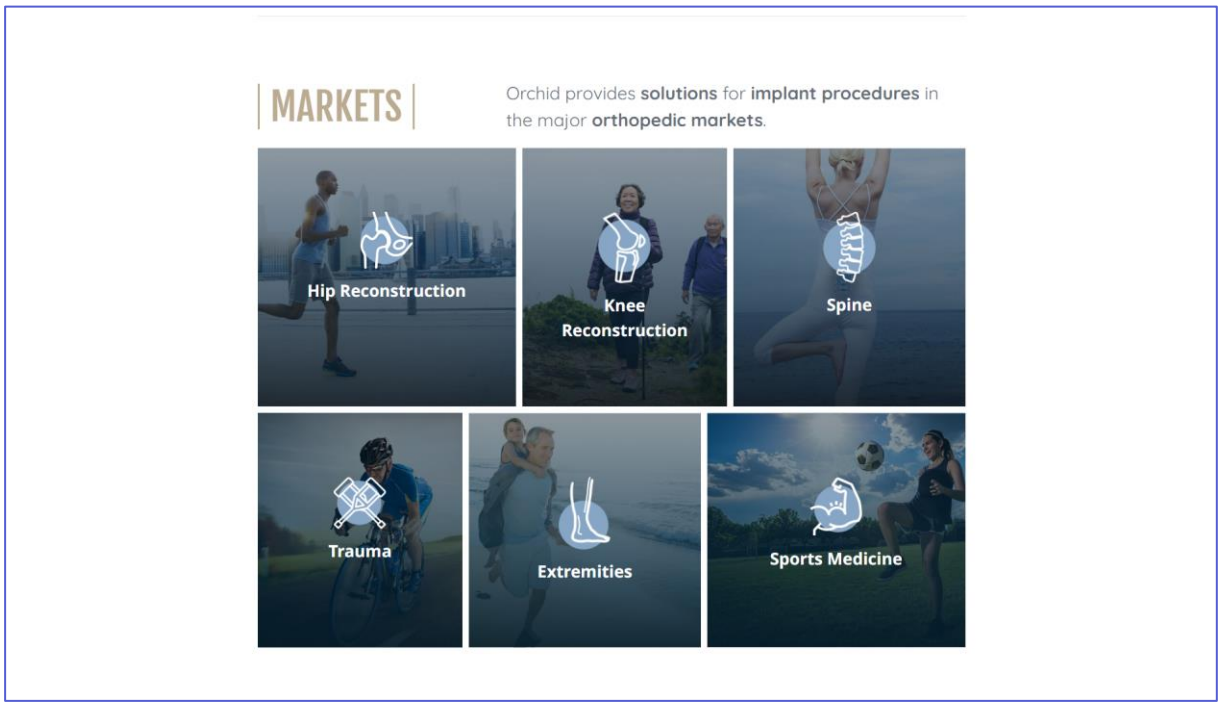
# Annex: Data Entity Examples

Supplier: Medtronic

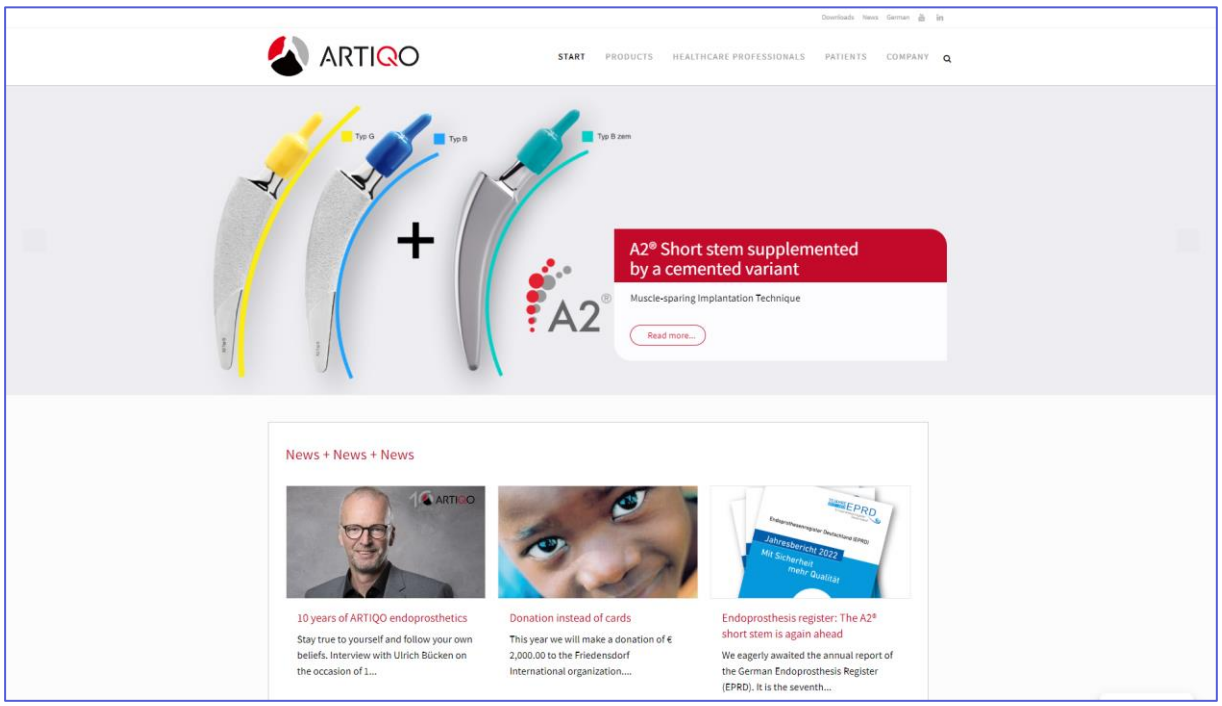


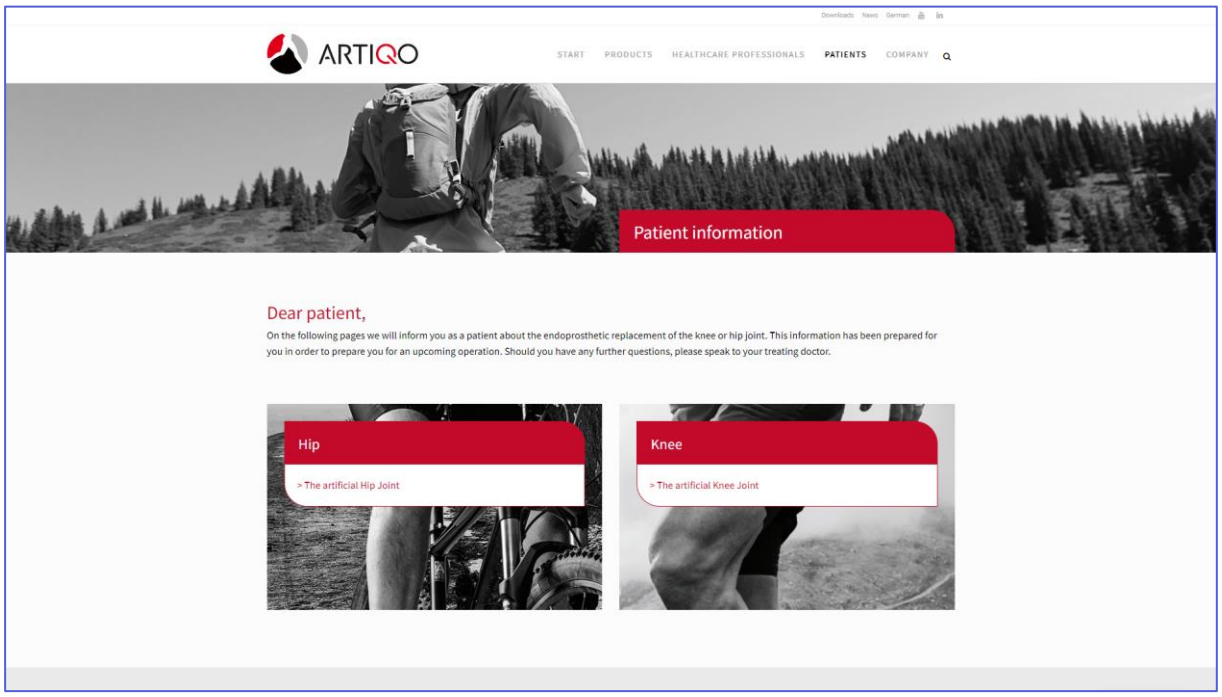



Supplier: Orchid Orthopedic Solutions

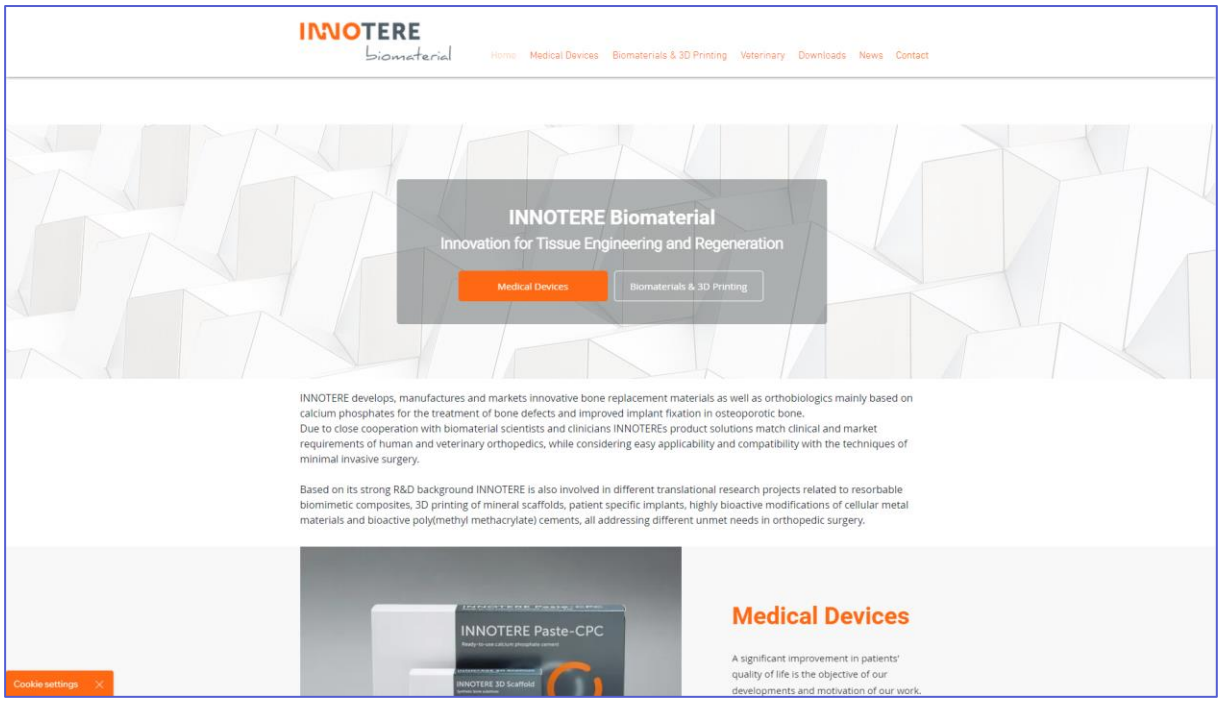



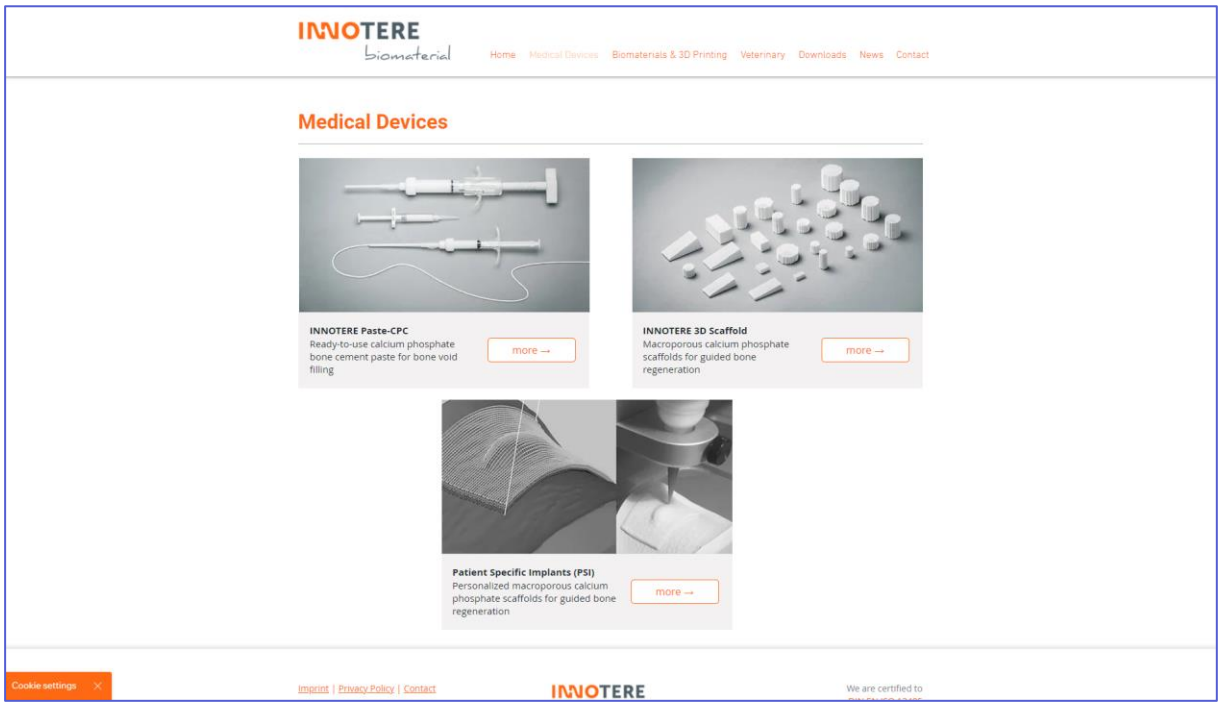
Supplier: ARTIQO



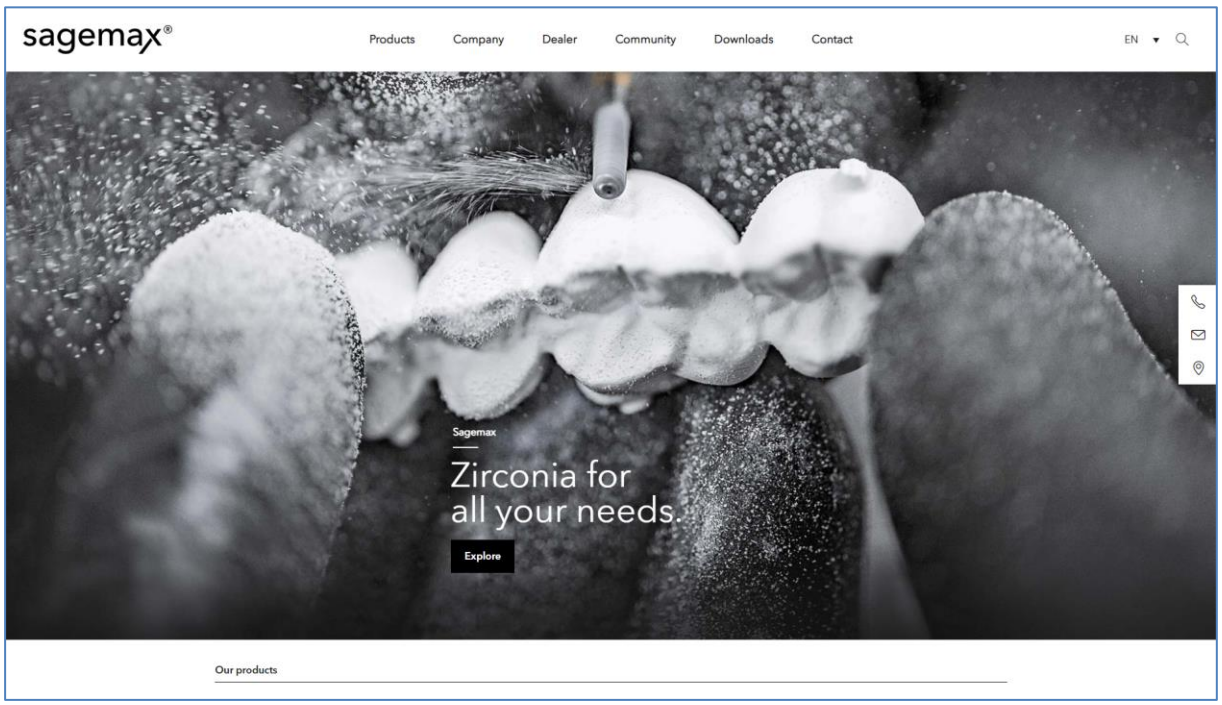



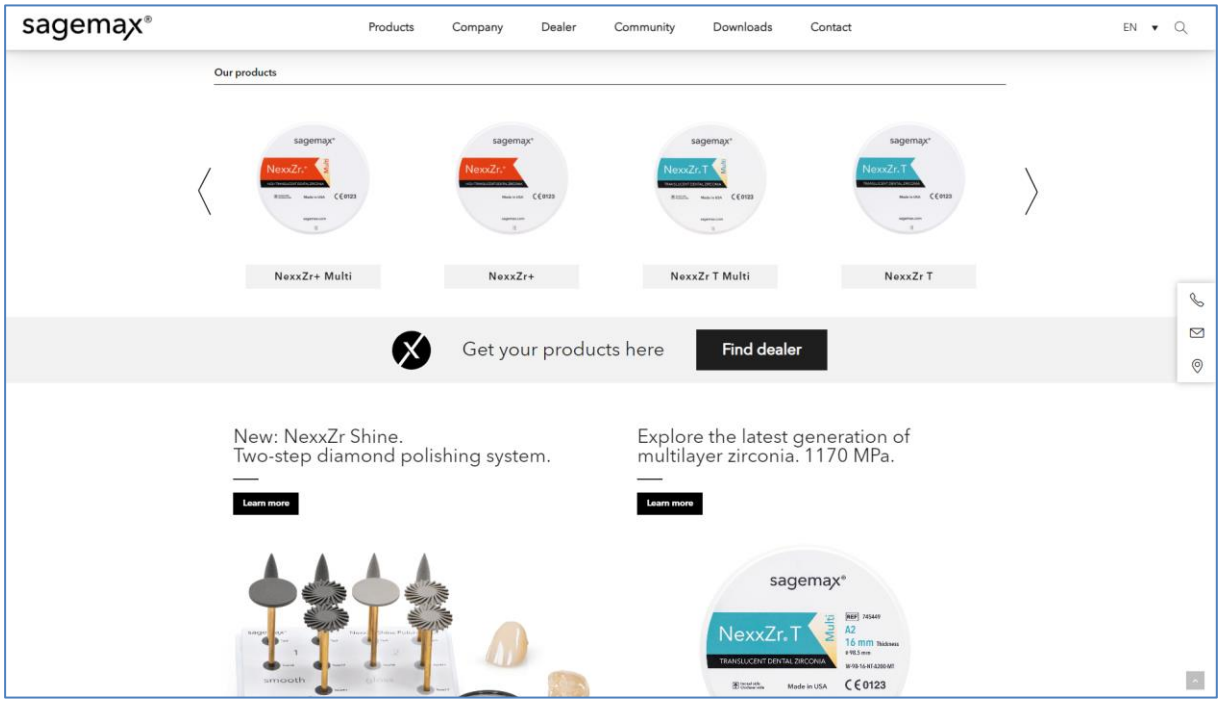
Supplier: INNOTERE



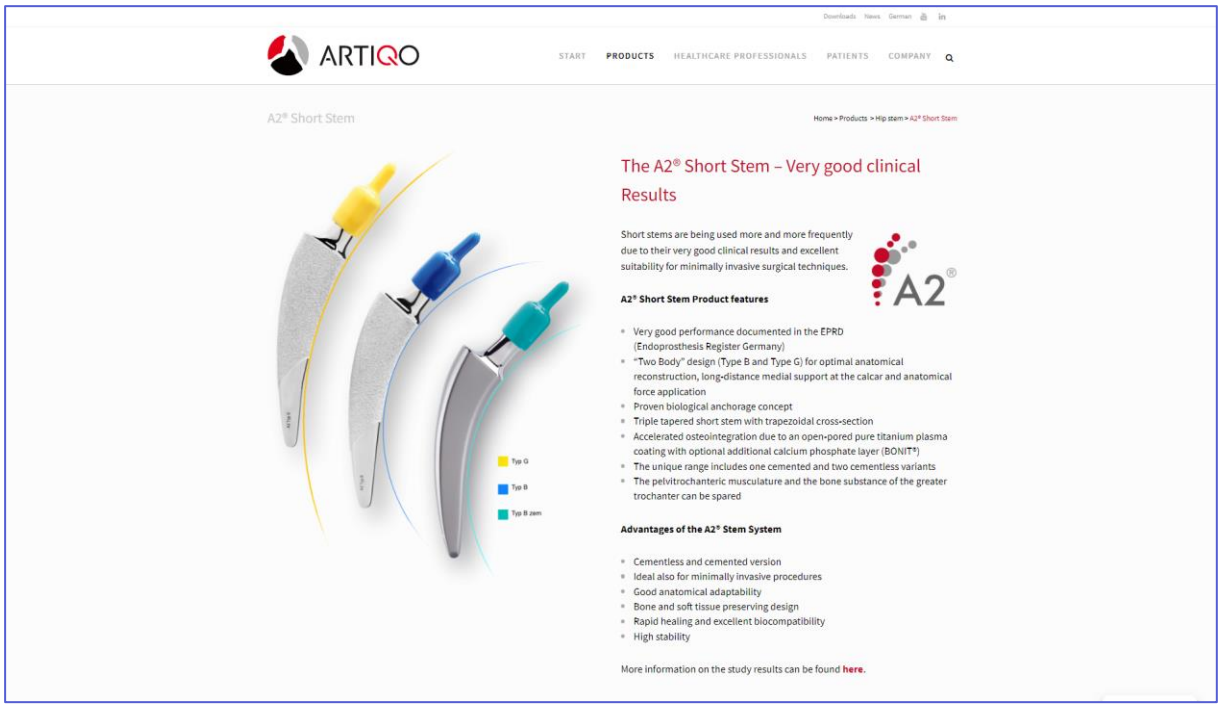



Supplier: Sagemax

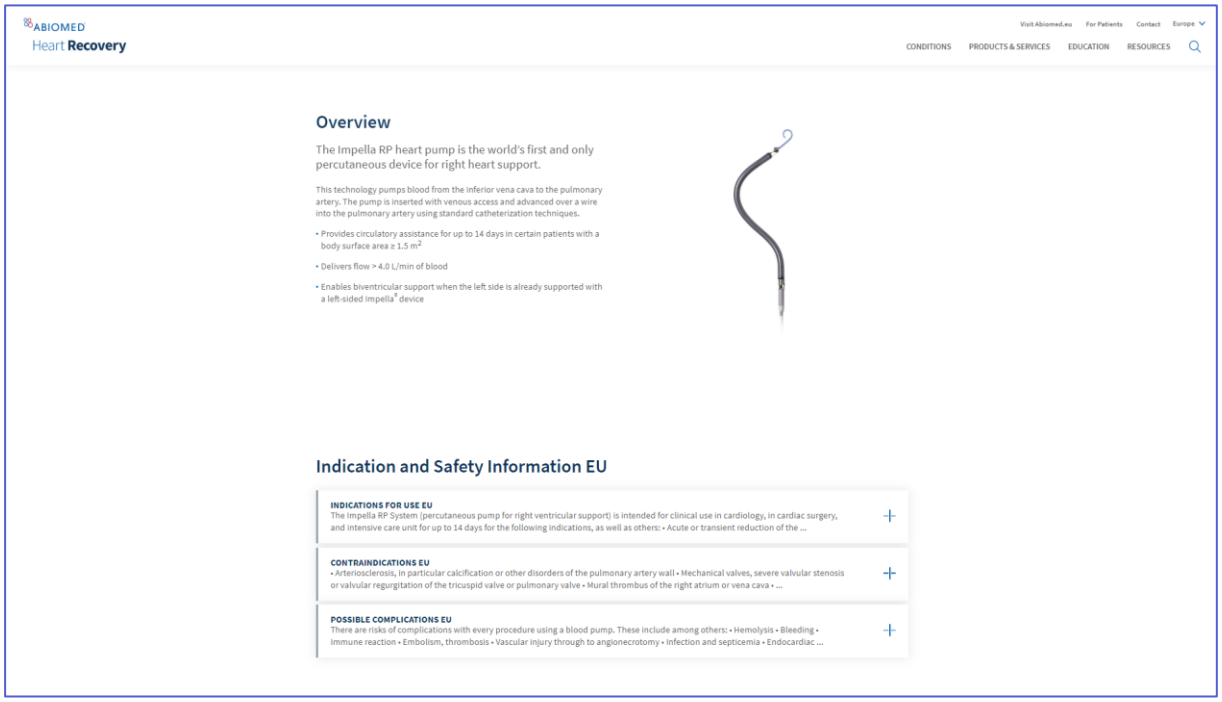


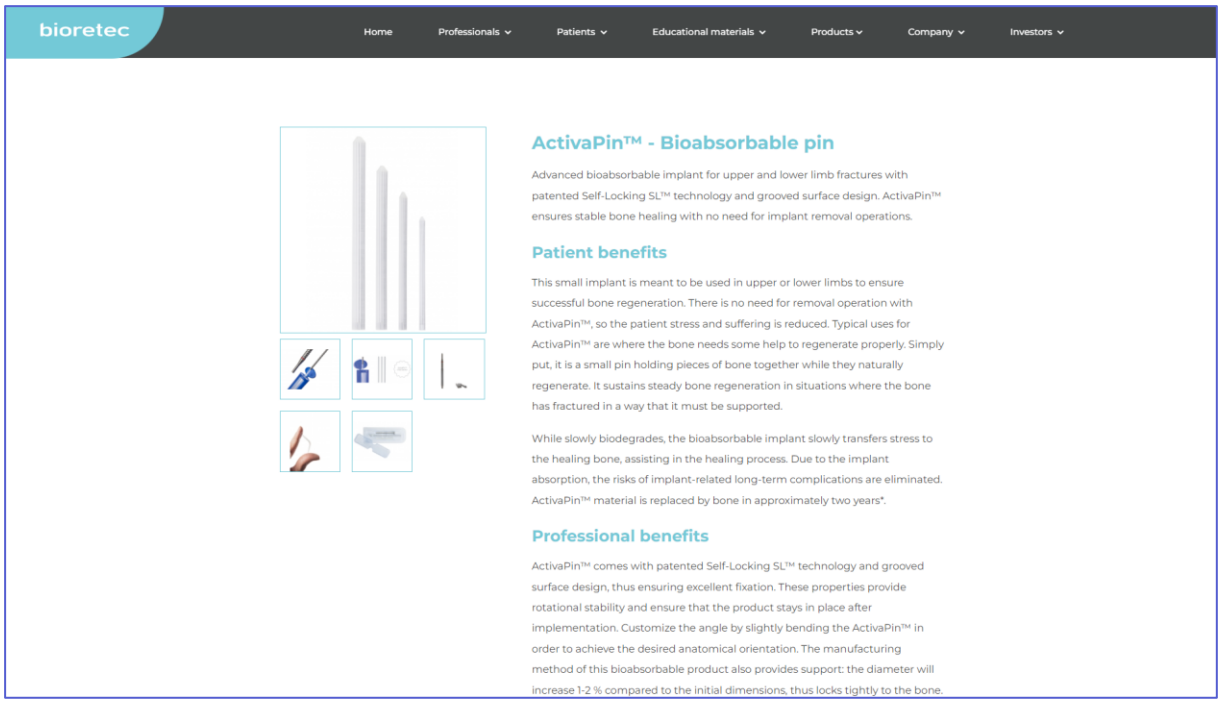
Product: A2 Short Stem



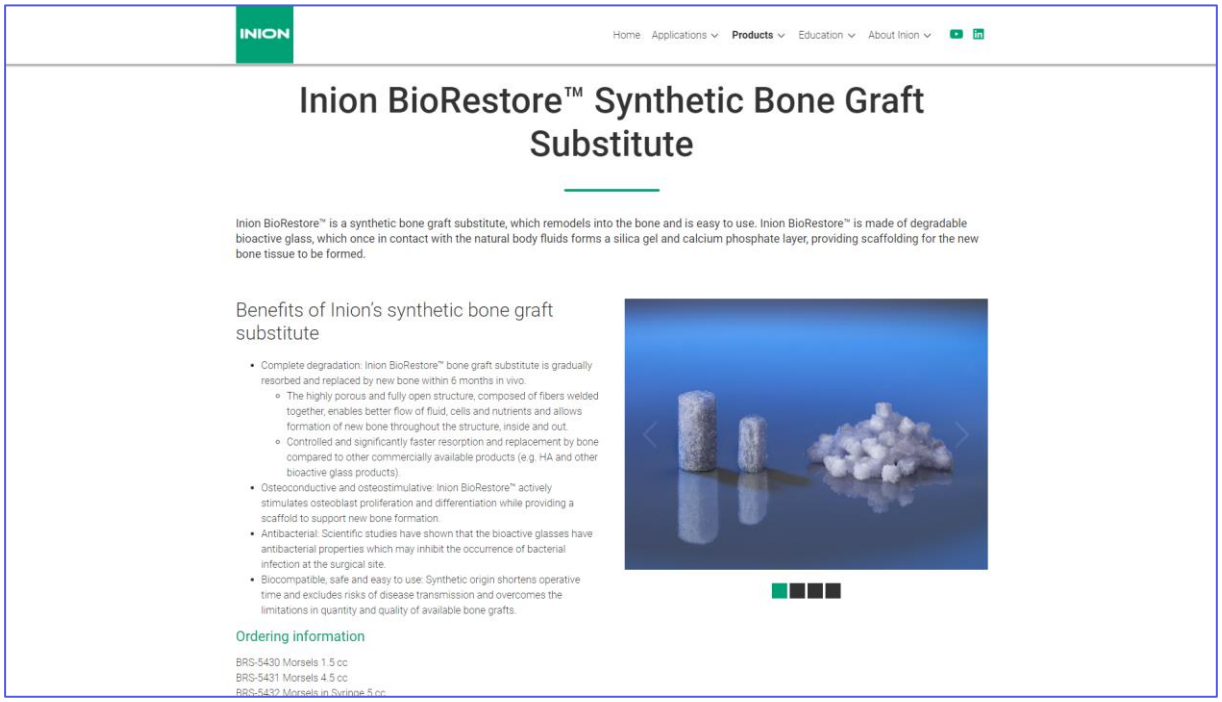

Product: Impella RP heart pump



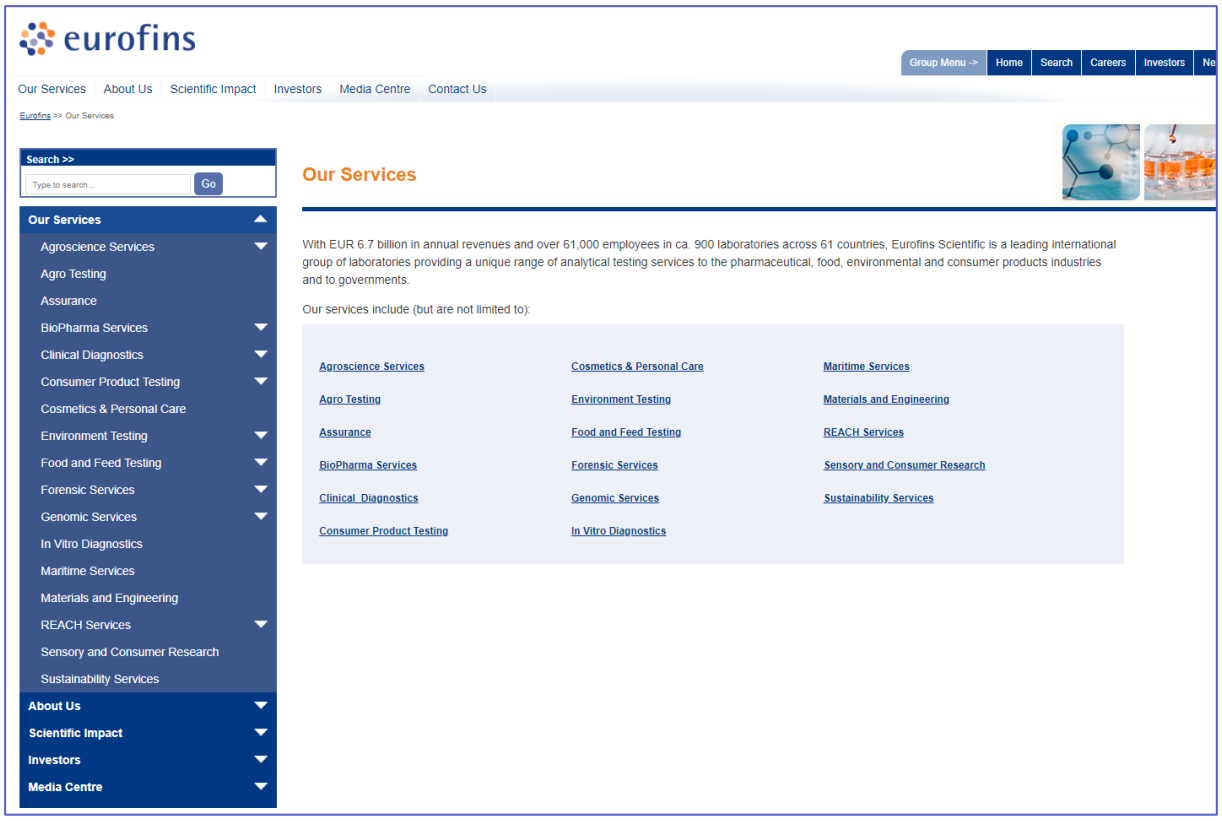
Product: ActivaPin



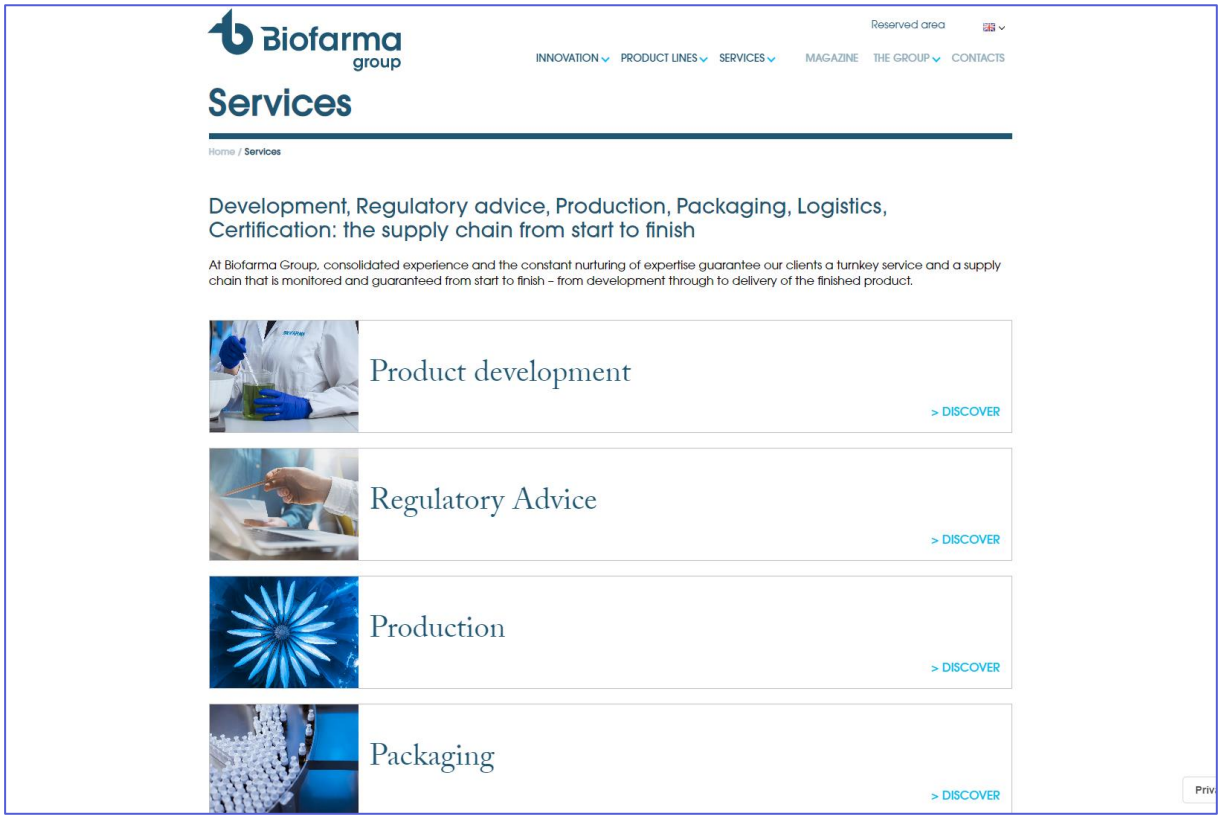

Product: Inion BioRestore



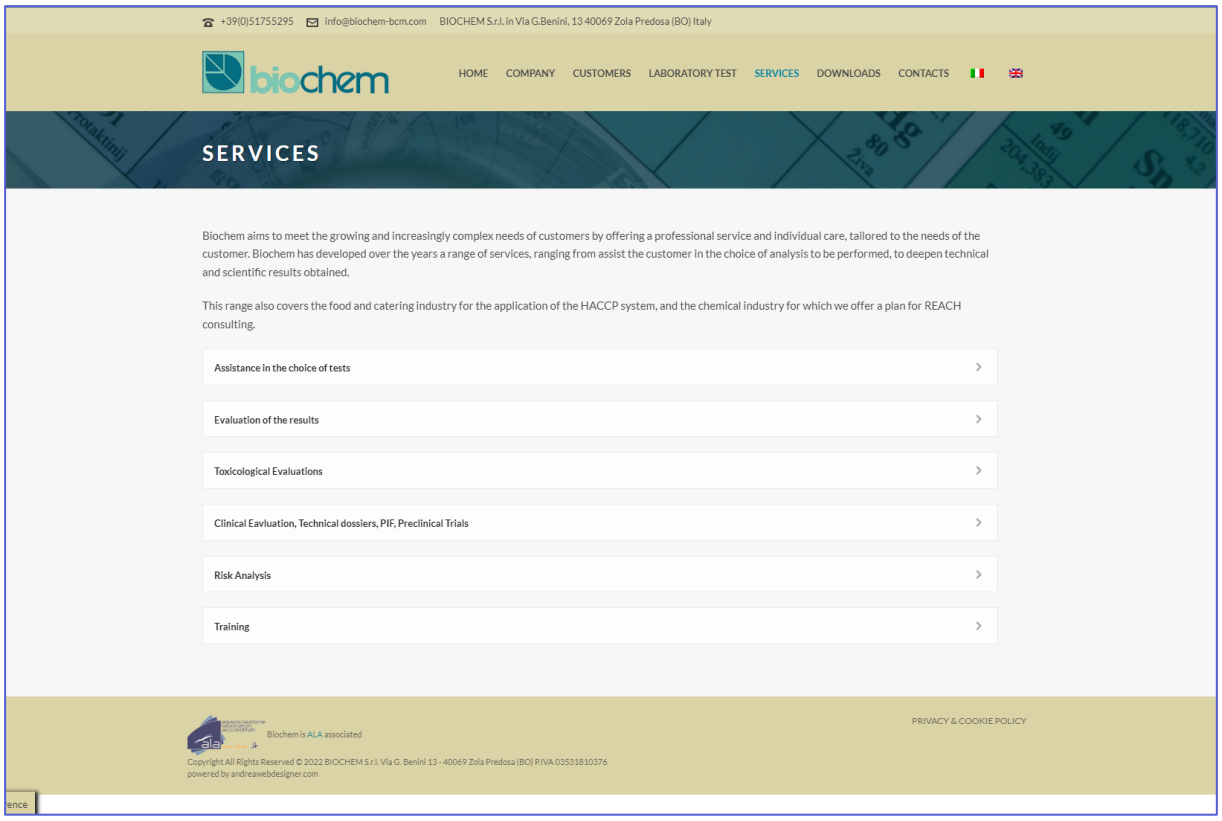
Services: Eurofins



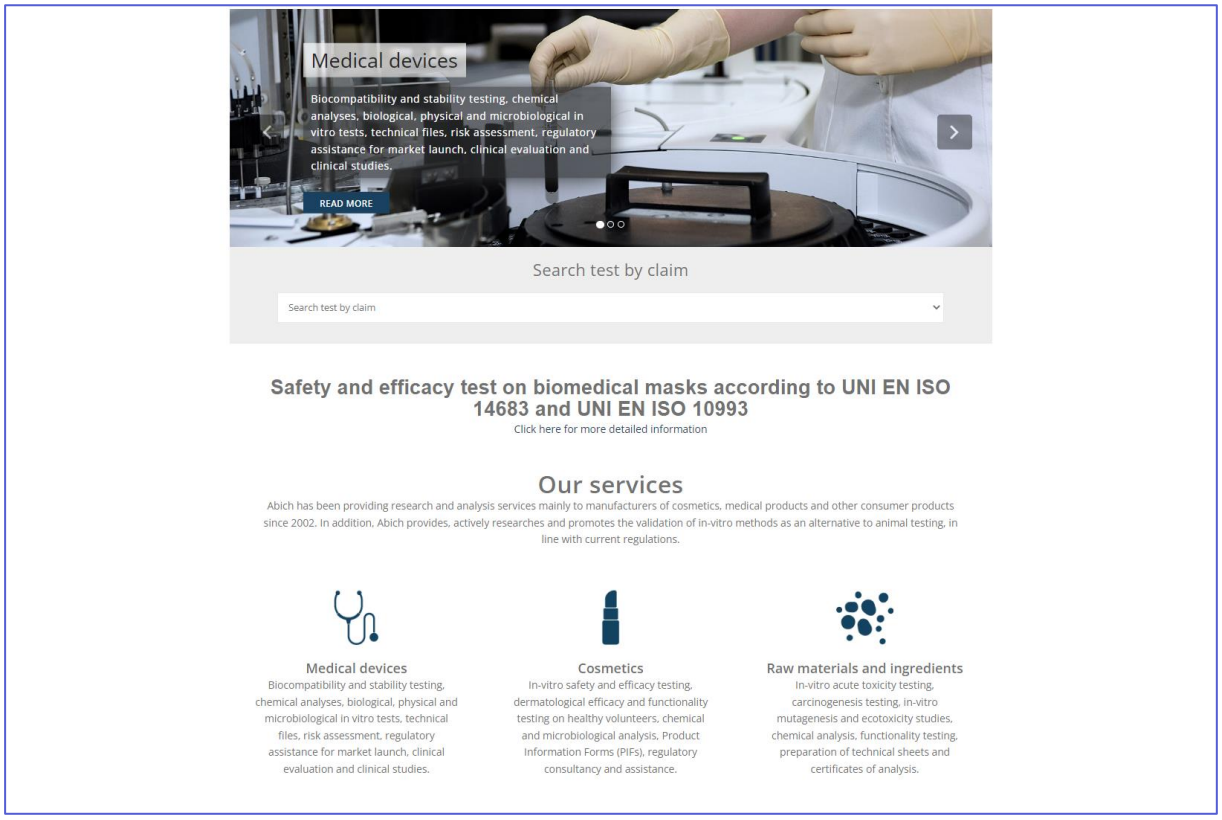

Services: Biofarma Group



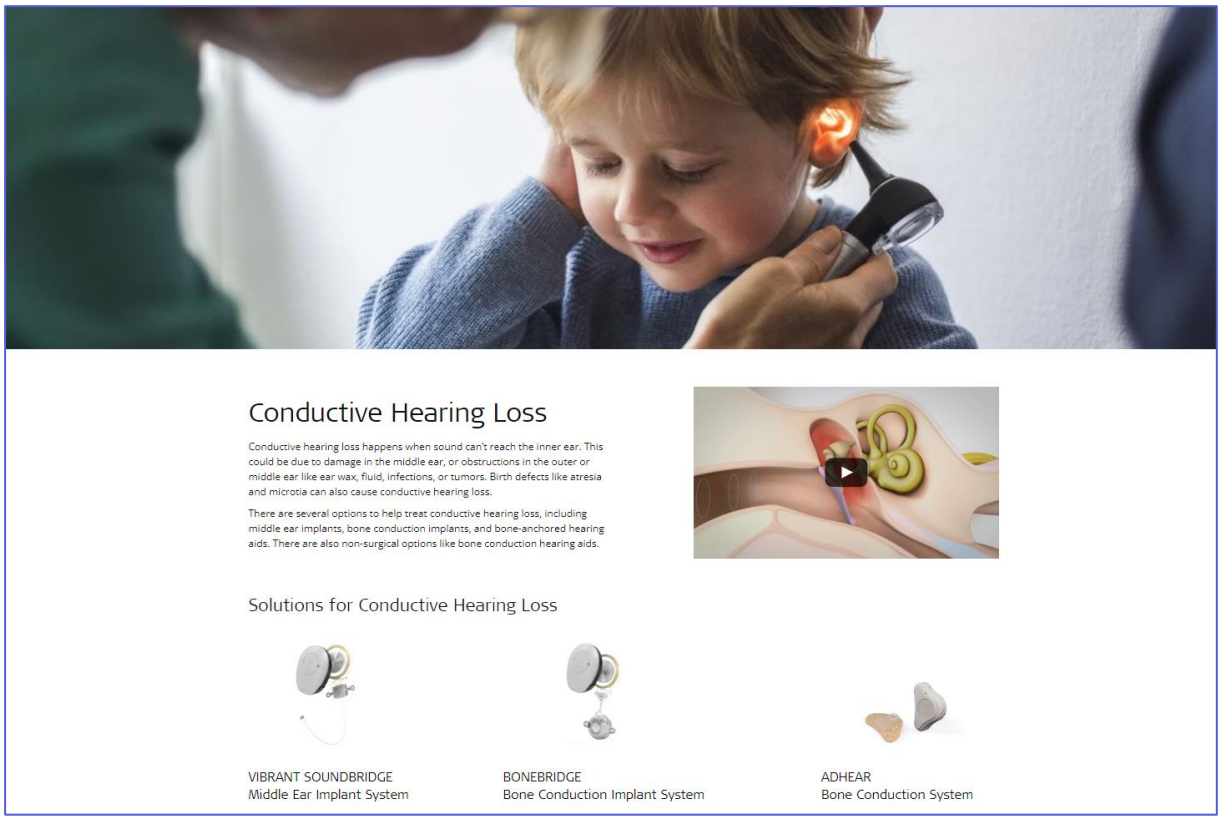
Services: Biochem

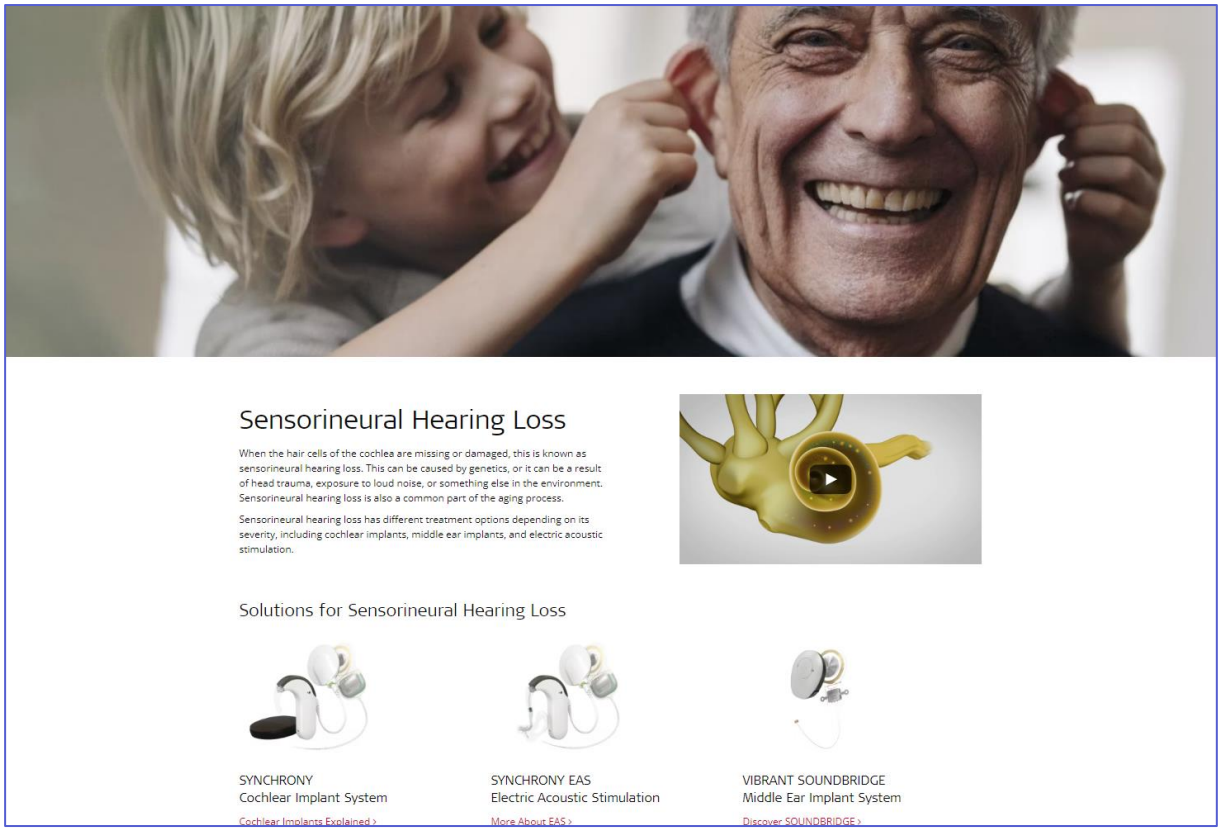



Services: Abich



Scenarios: MED-EL Types of Hearing Loss



## Mixed Hearing Loss

Mixed hearing loss is a combination of both sensorineural and conductive hearing loss. It results from problems in both the inner and the outer or middle ear.

Options for treating mixed hearing loss include middle ear implants and bone conduction implants.

### Solutions for Mixed Hearing Loss

VIBRANT SOUNDBRIDGE
Middle Ear Implant System

BONEBRIDGE
Bone Conduction Implant

## Neural Hearing Loss

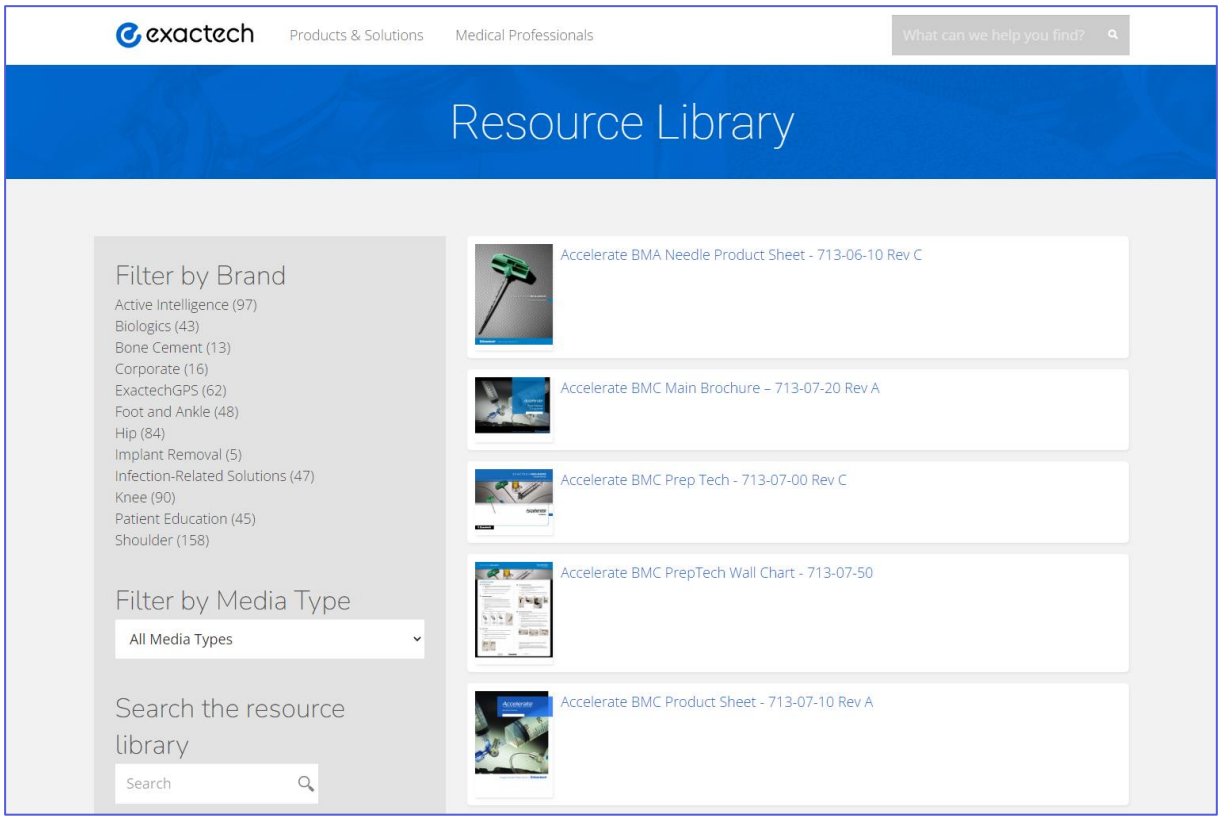
When the auditory nerve is damaged or missing, this is known as neural hearing loss. Hearing aids and cochlear implants cannot help because the nerve is not able to pass on sound information to the brain.

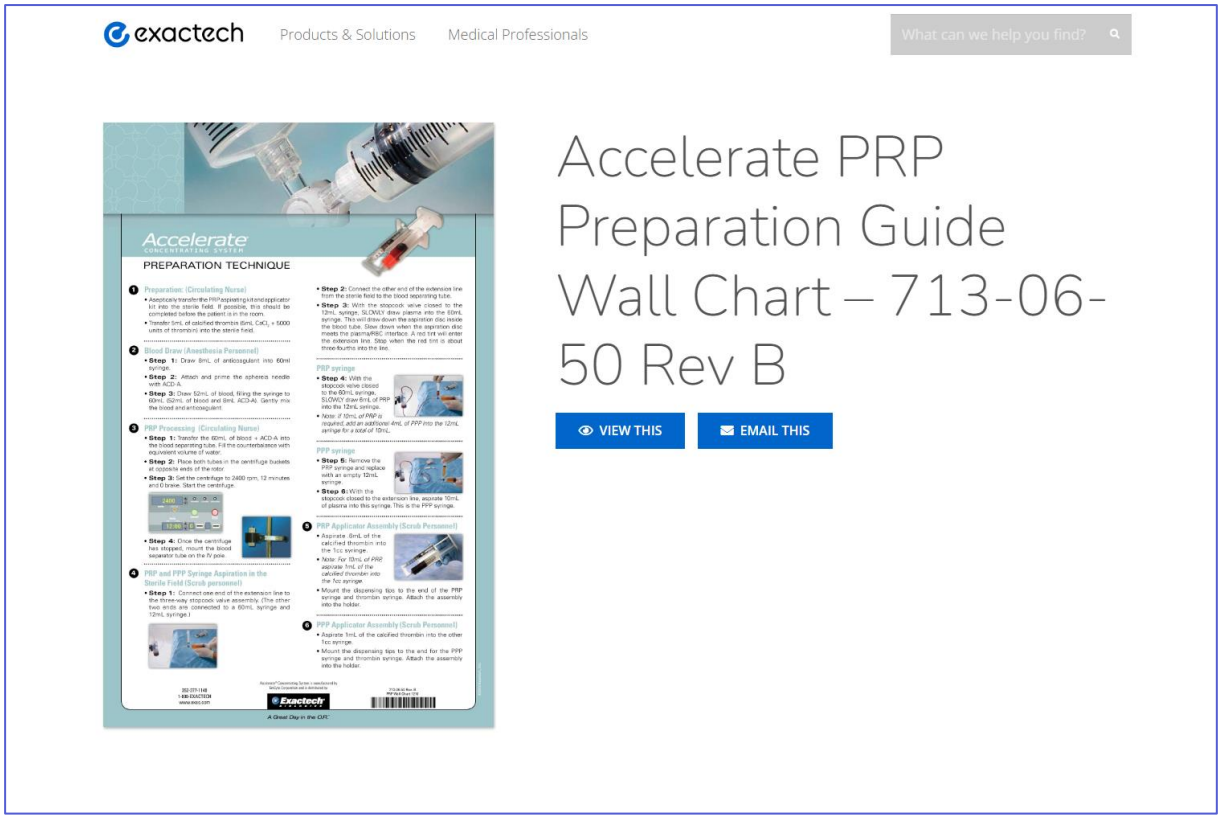
In many cases, an Auditory Brainstem Implant (ABI) is an option for neural hearing loss.

### Solutions for Neural Hearing Loss

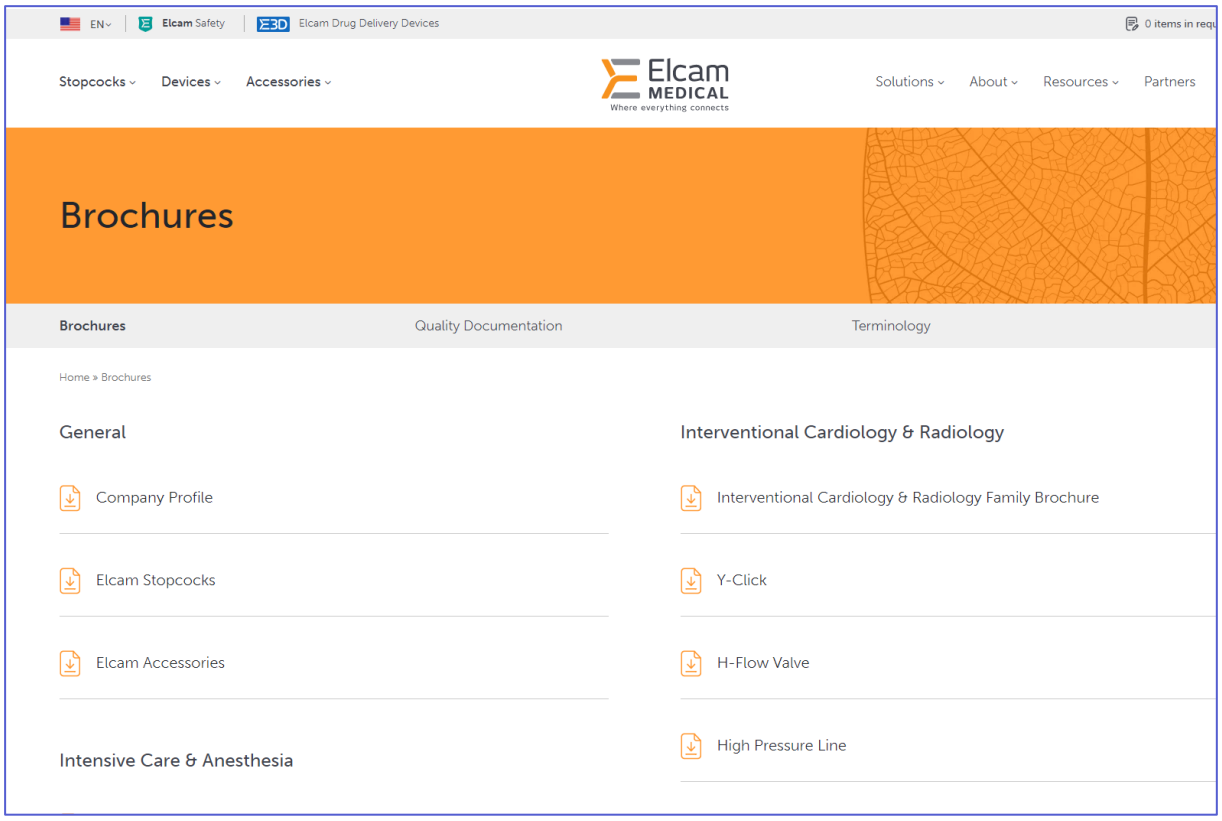
SYNCHRONY ABI
Auditory Brainstem Implant System

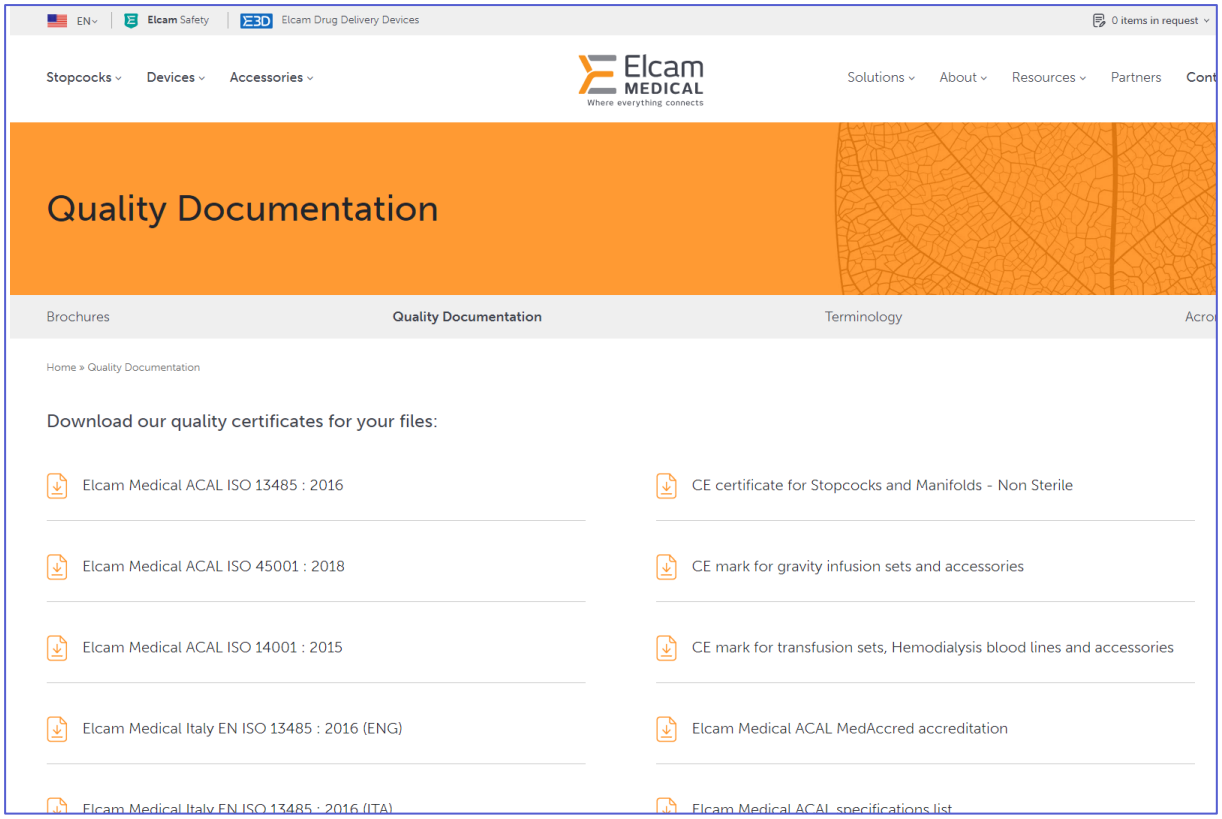
Documents: Exactech Resource Library



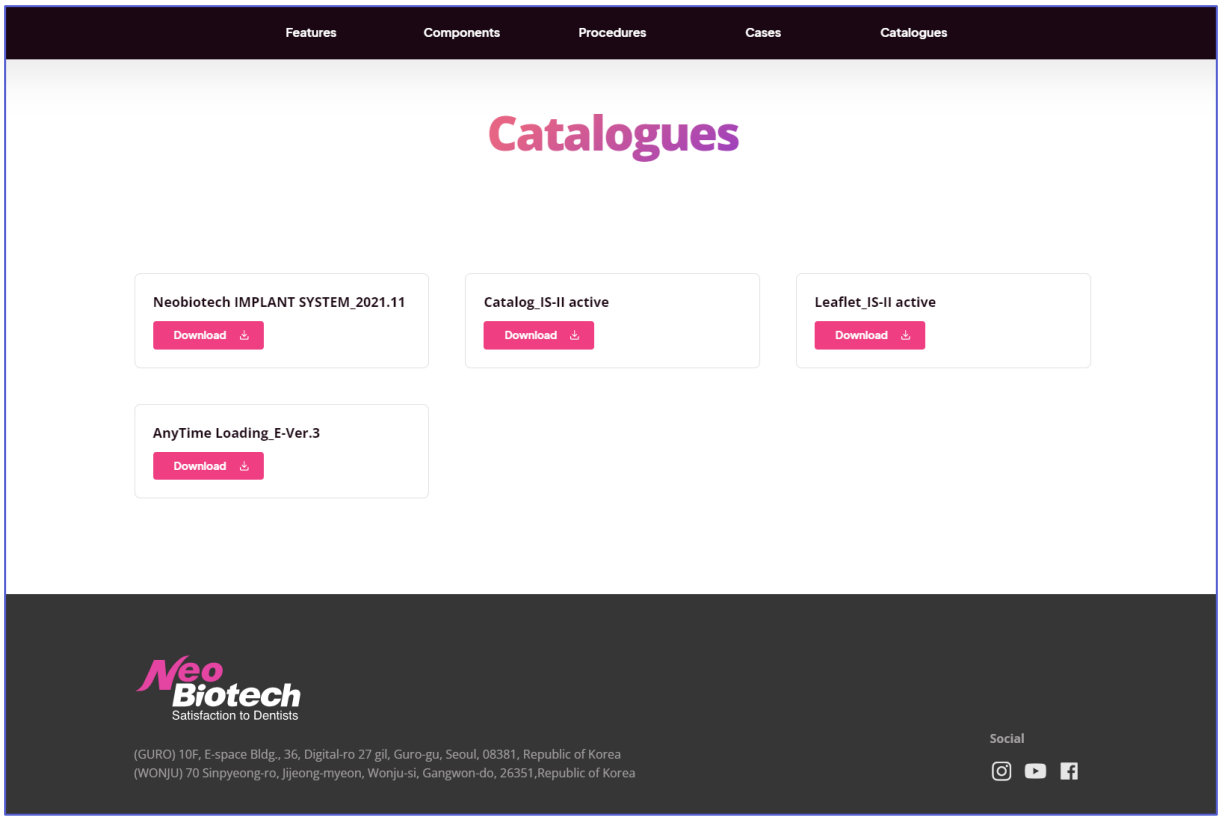



Documents: Elcam Medical Brochures & Documentations

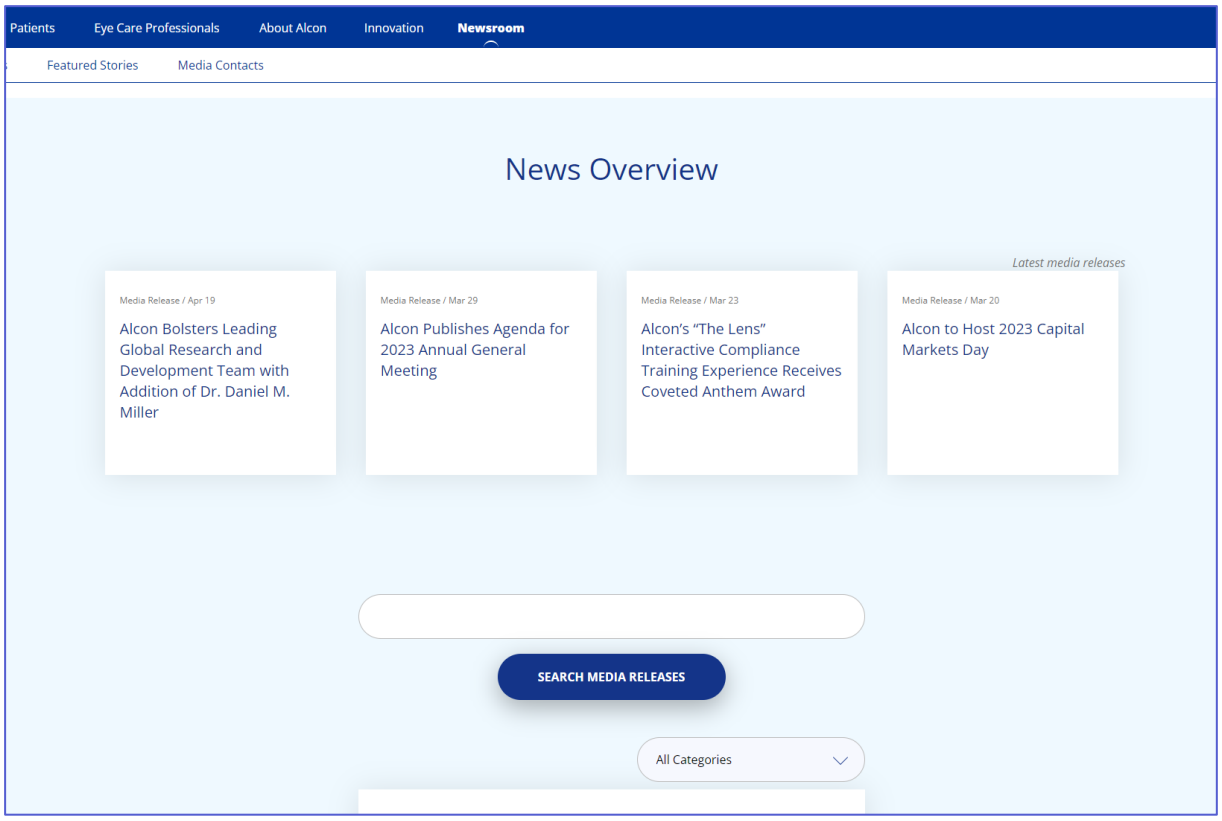


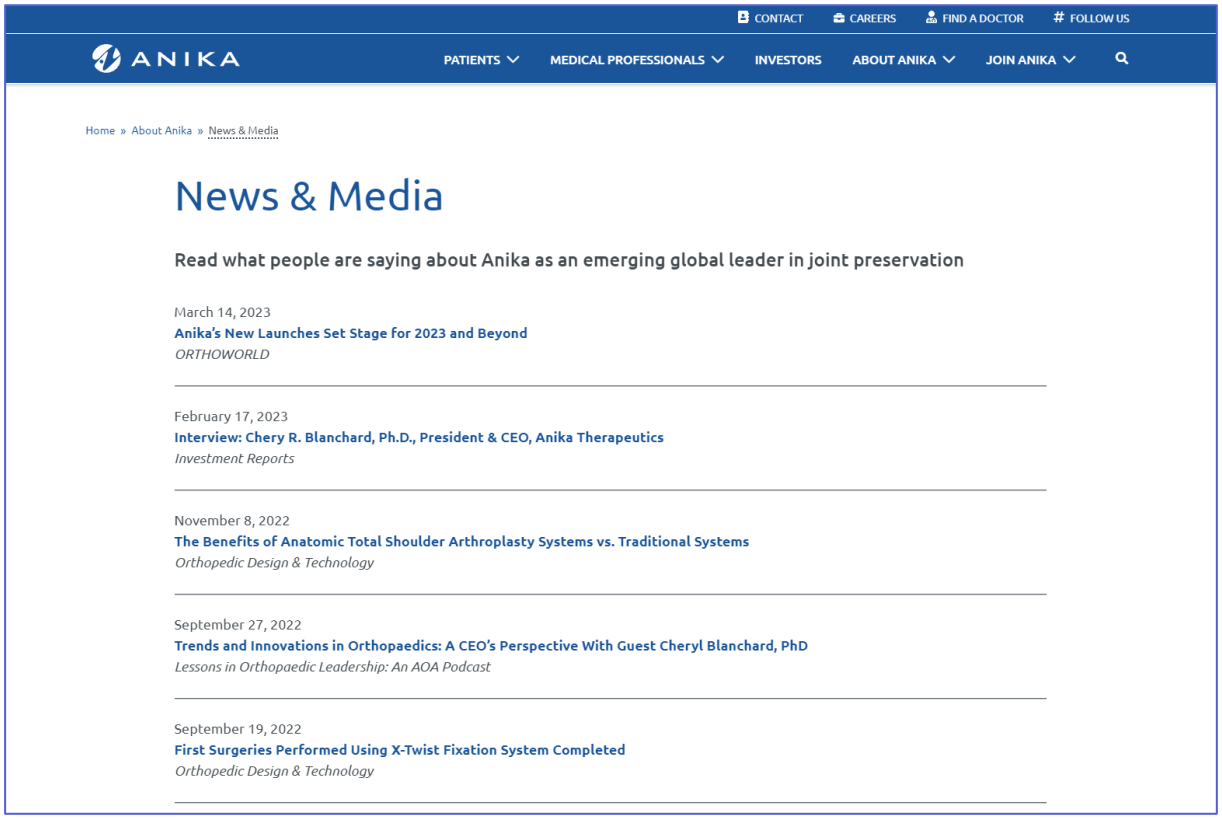
Documents: Neobiotech Product Catalogues



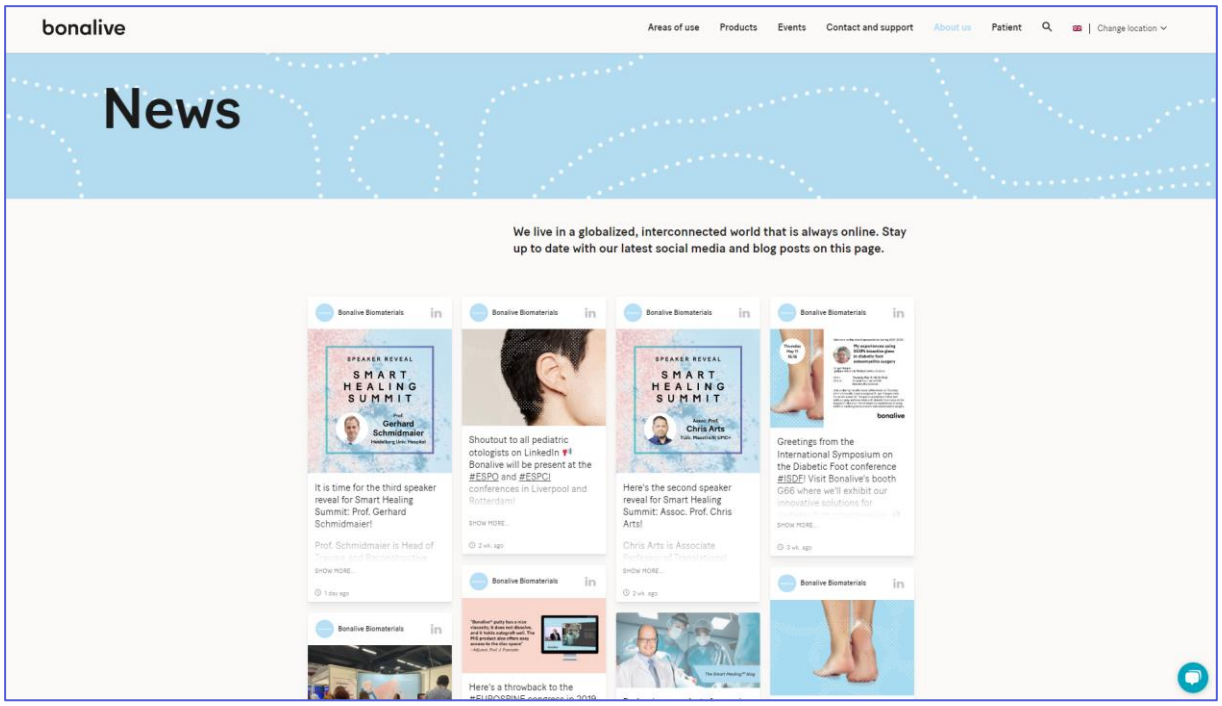

News: Alcon Newsroom



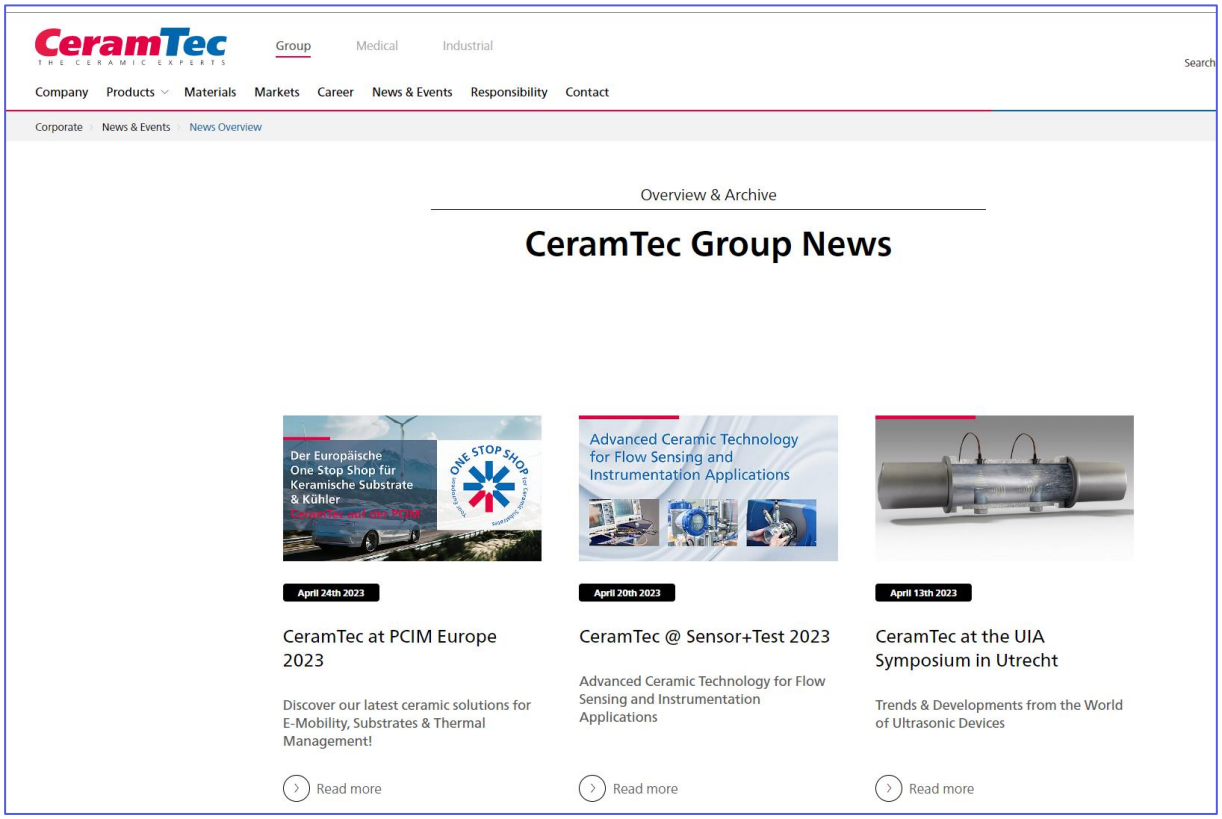

News: Anika



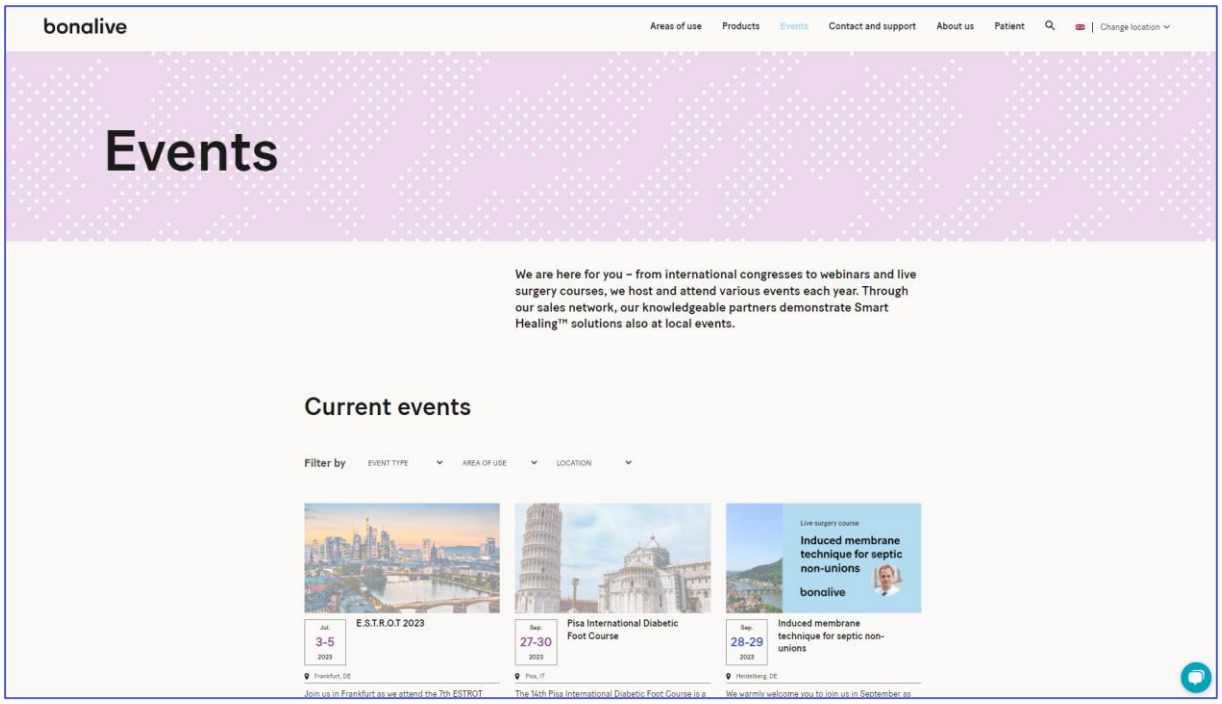
News: Bonalive



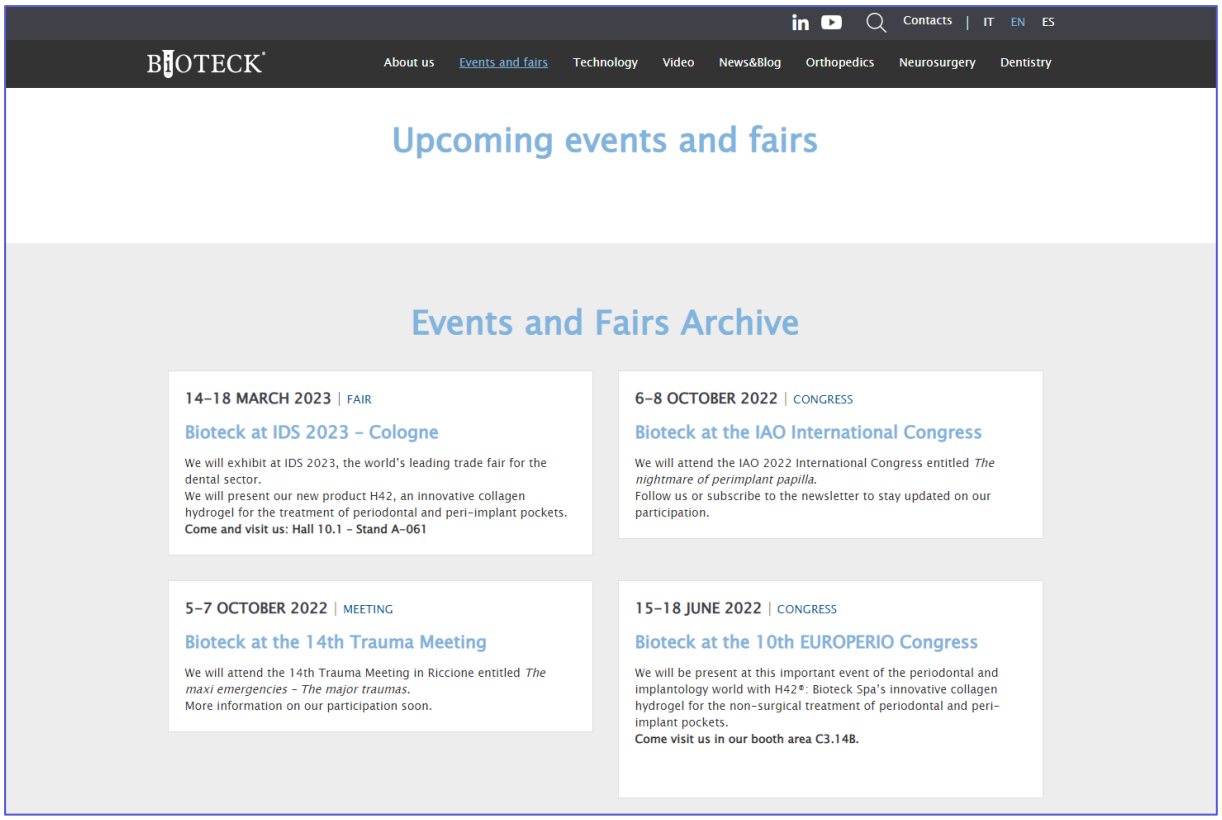

News: CeramTec



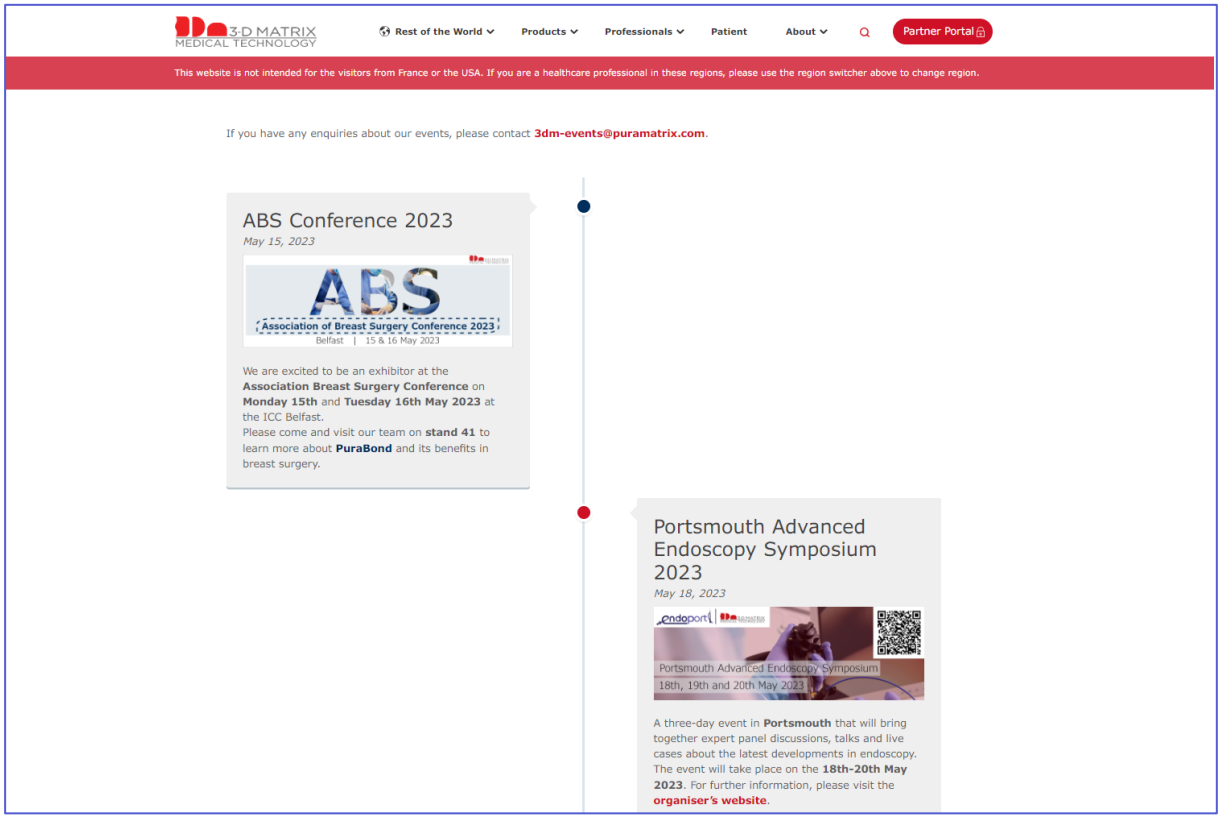
Events: Bonalive



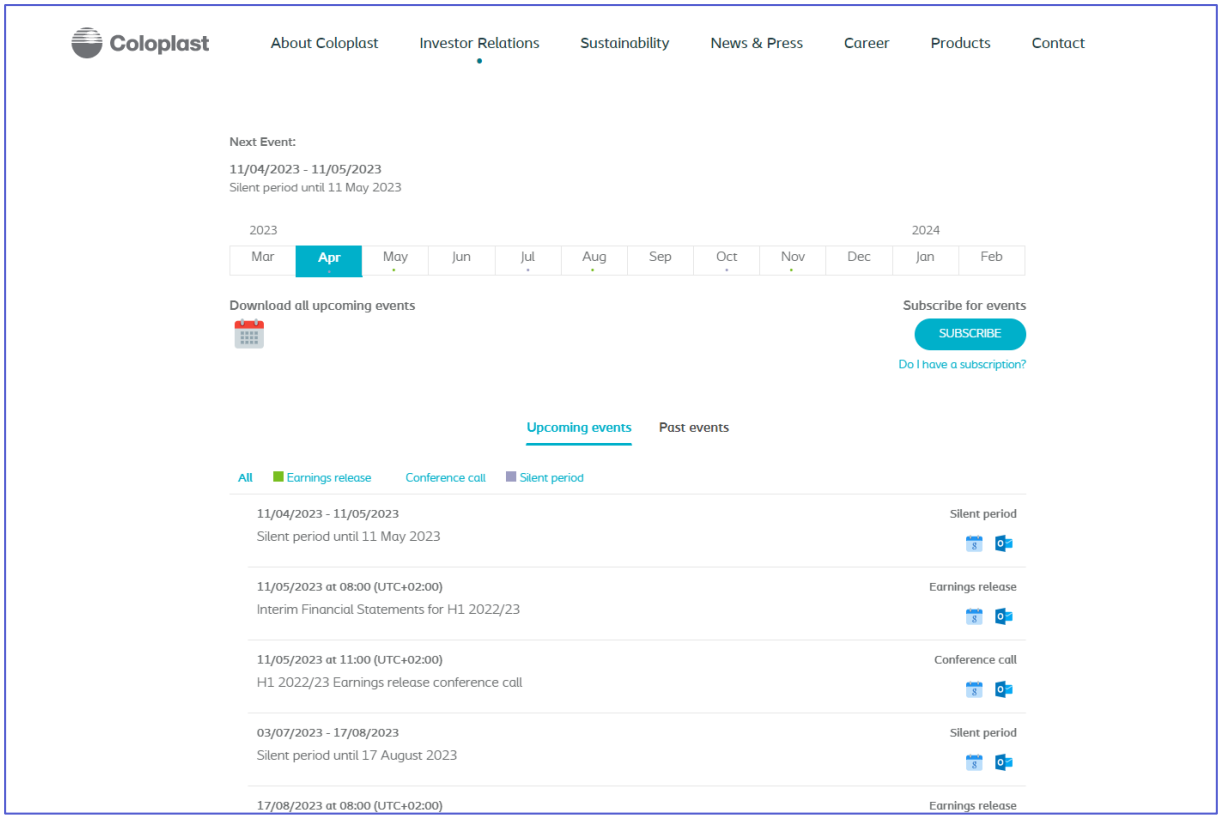

Events: Bioteck



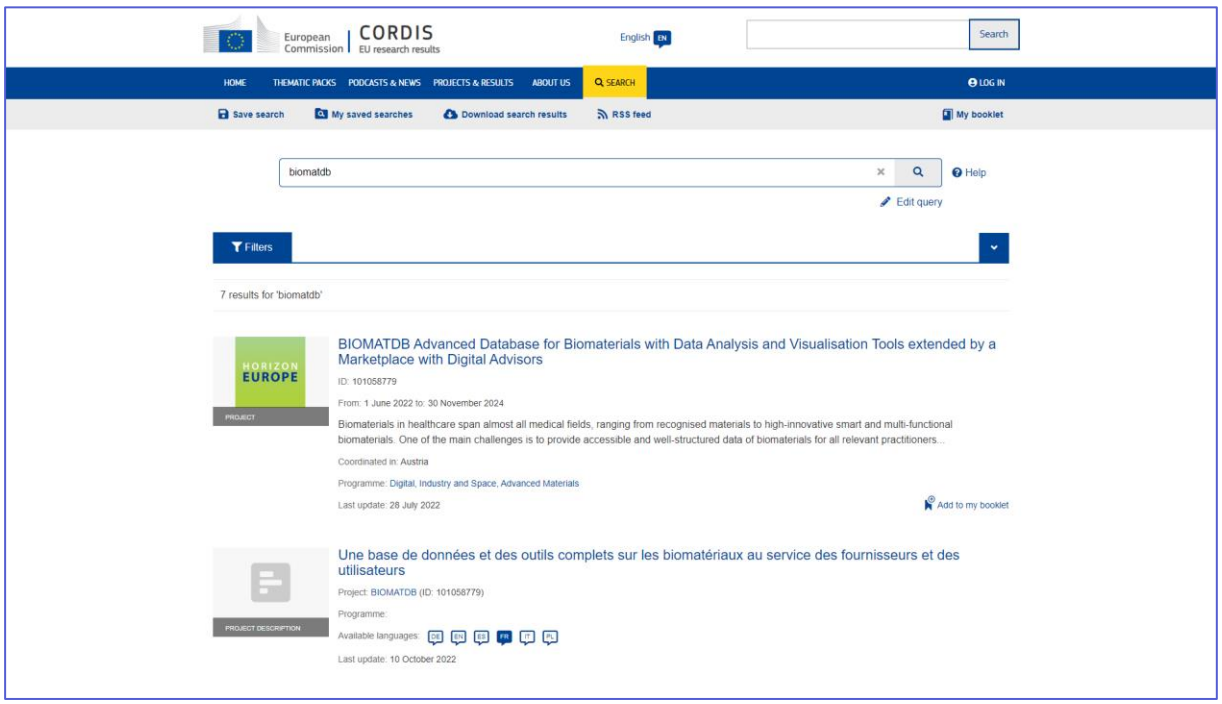
Events: 3-D Matrix

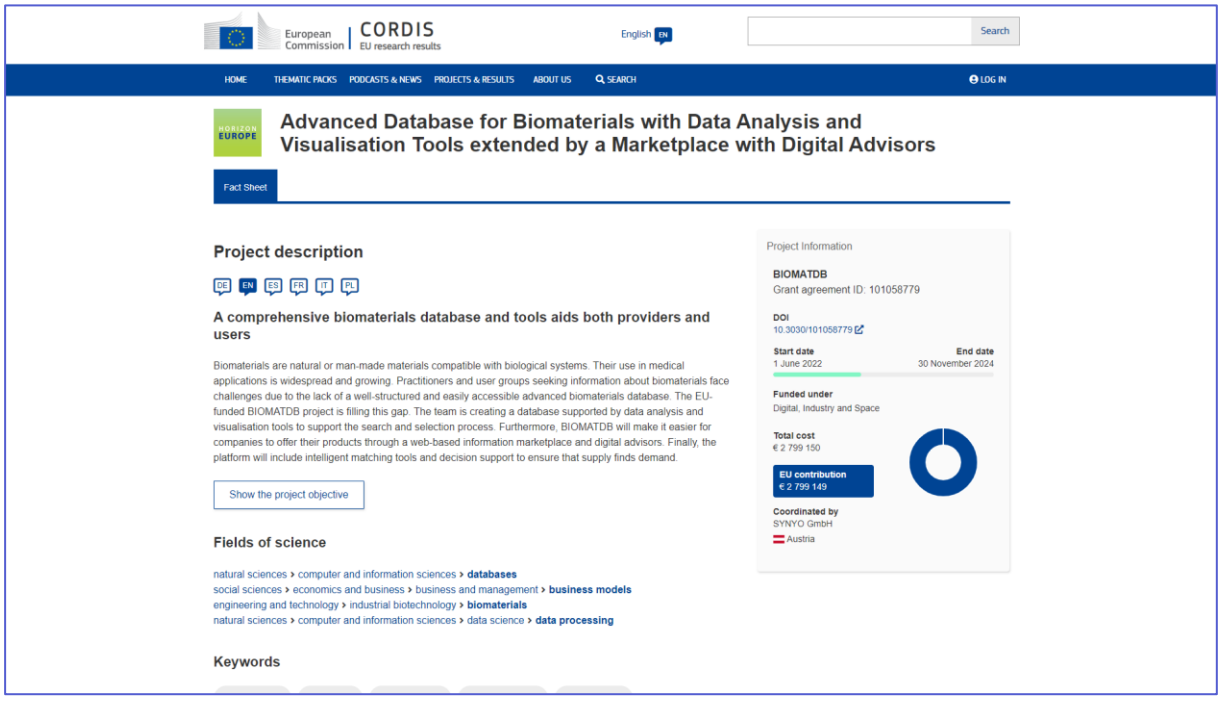


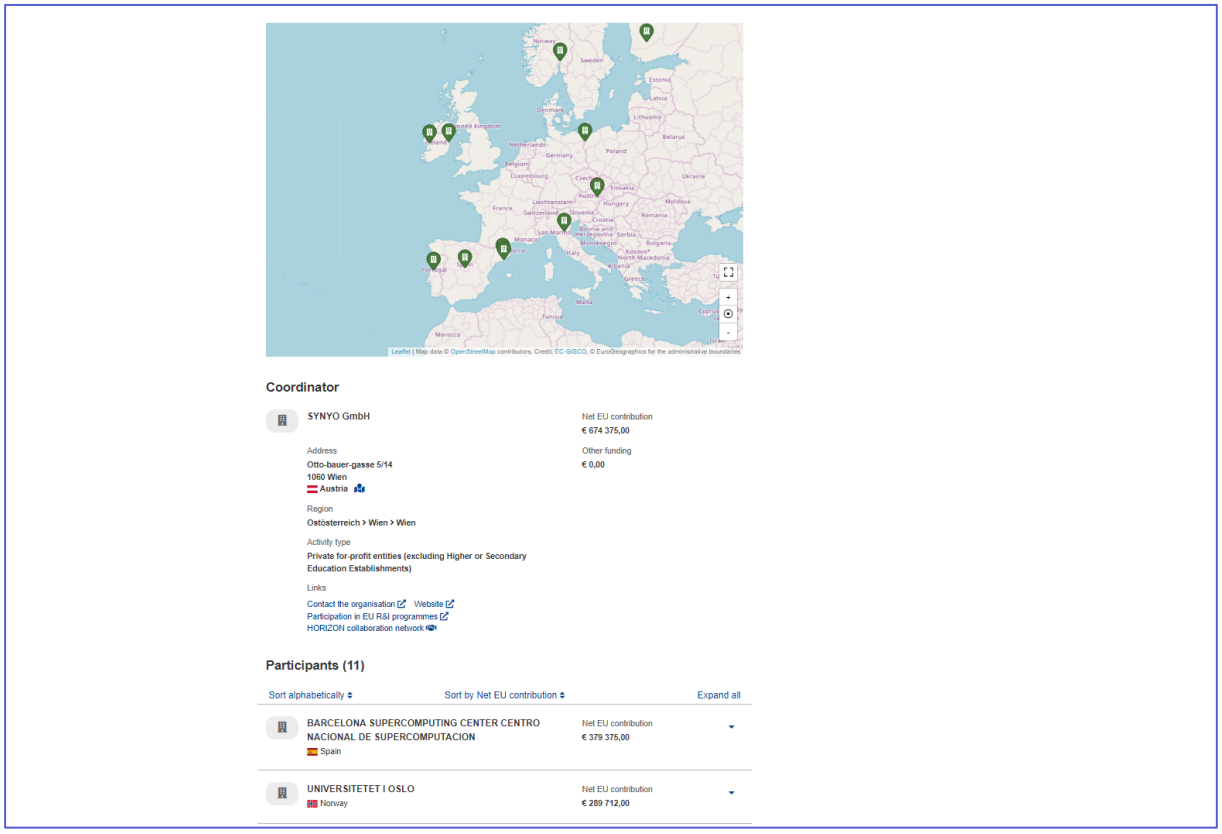

Events: Coloplast



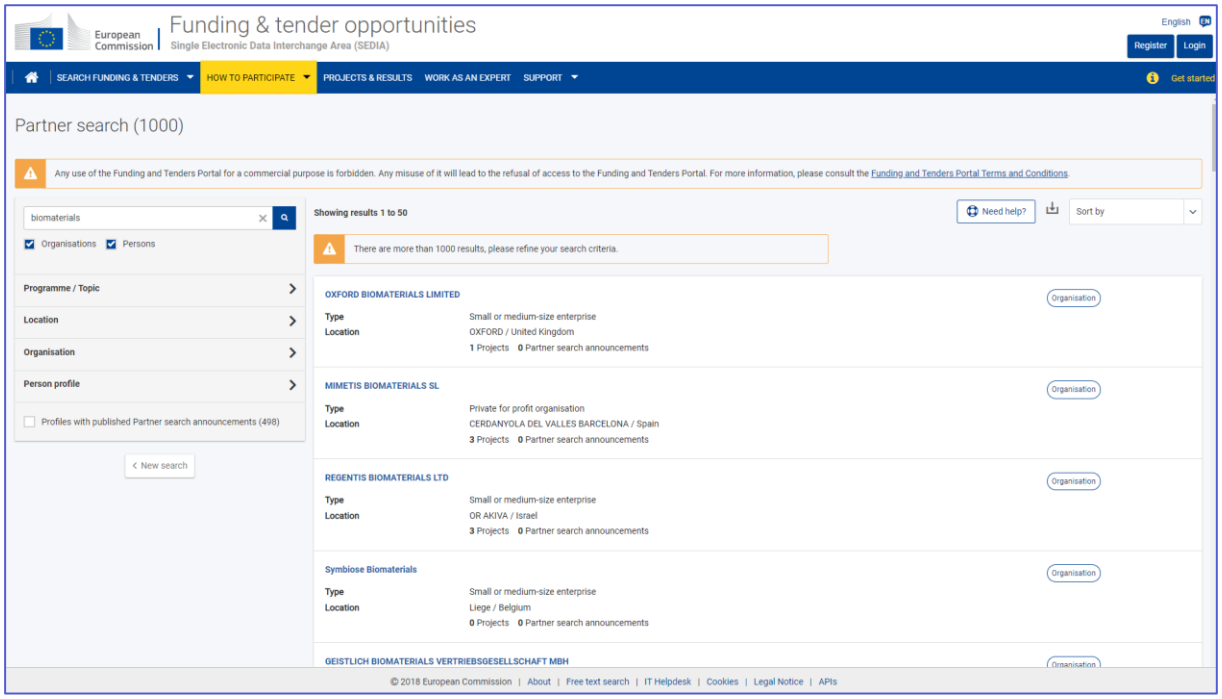
EU Projects: CORDIS

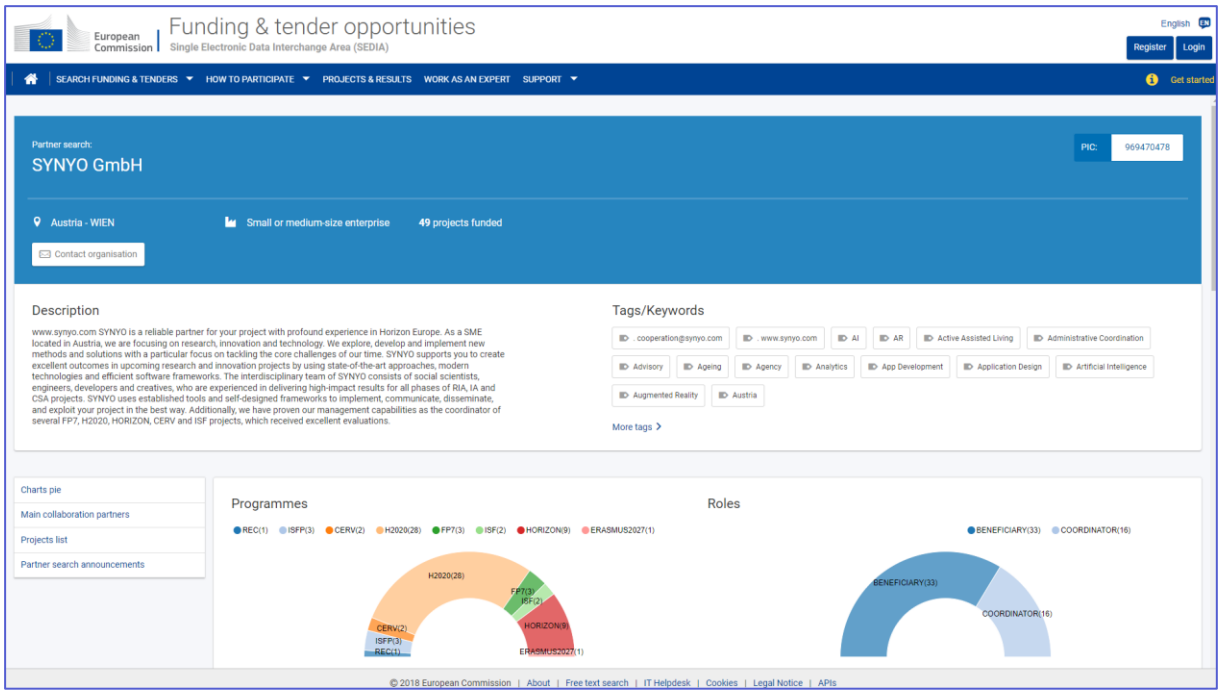


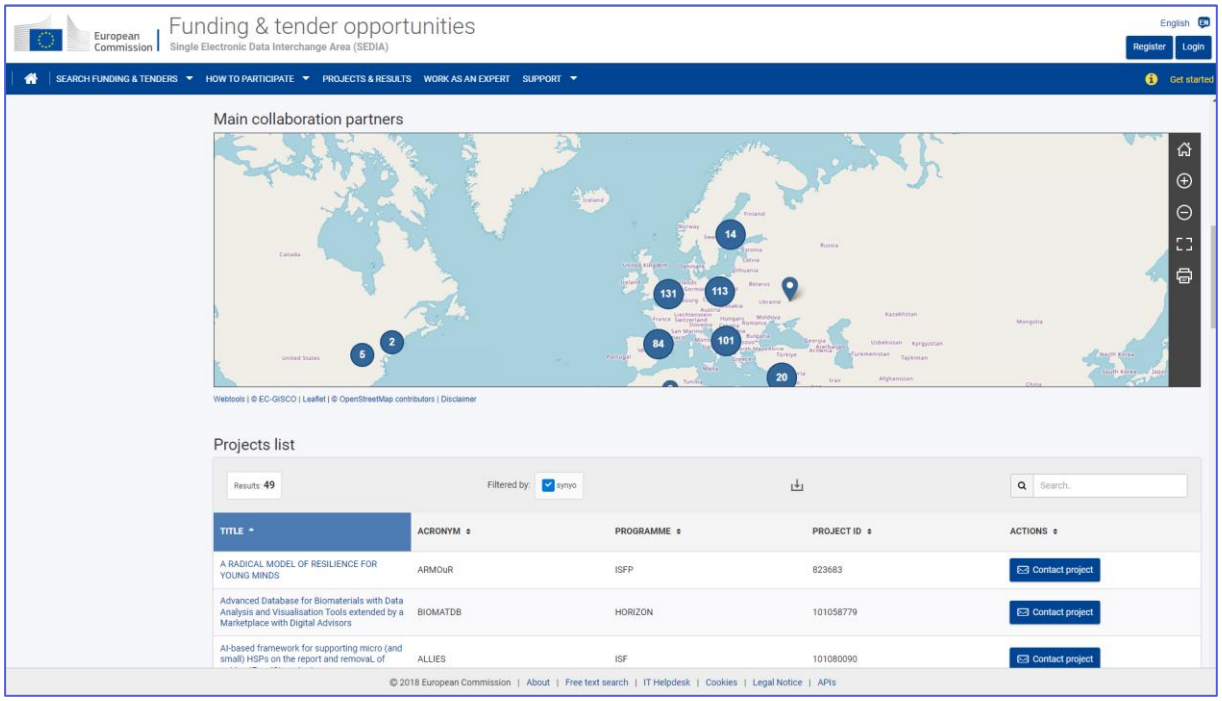





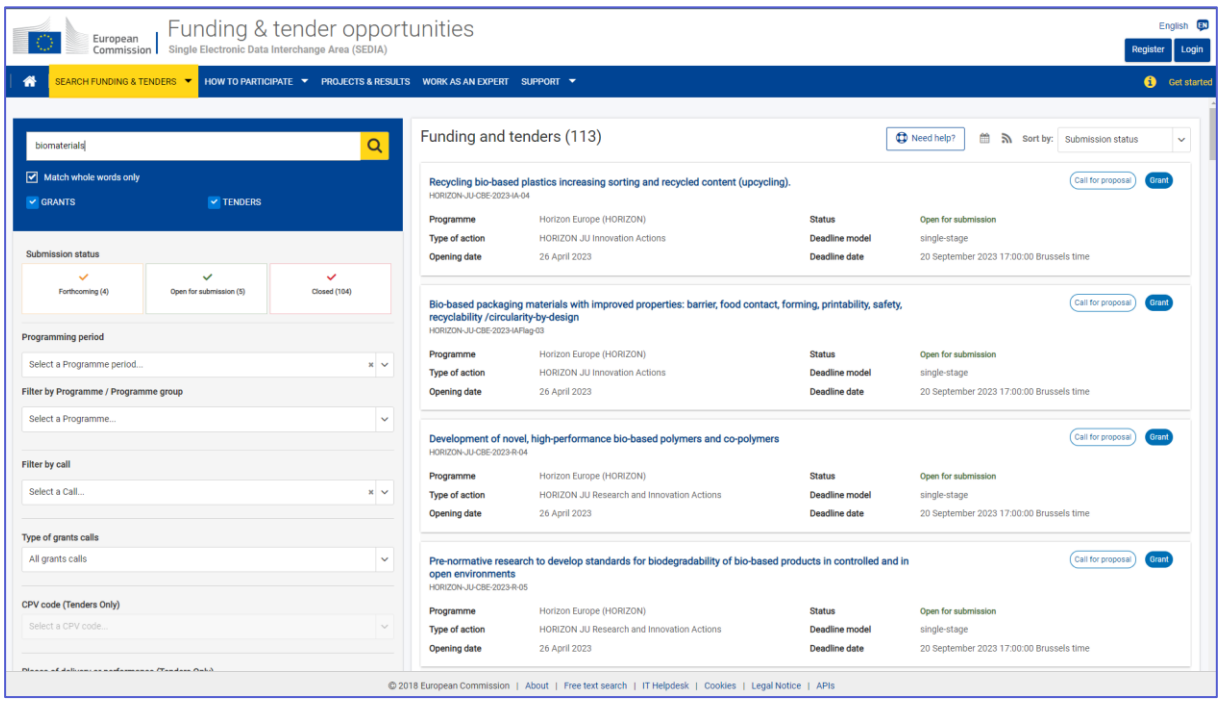

EU Project Partners: EC Funding and Tender Opportunities

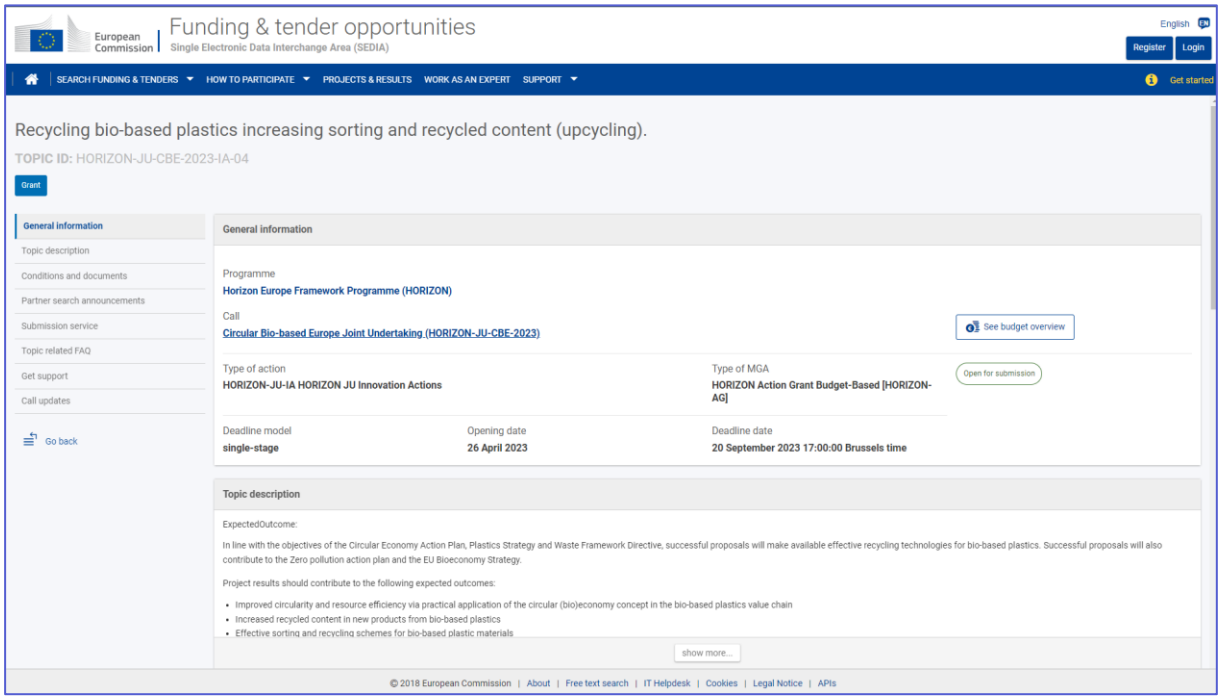


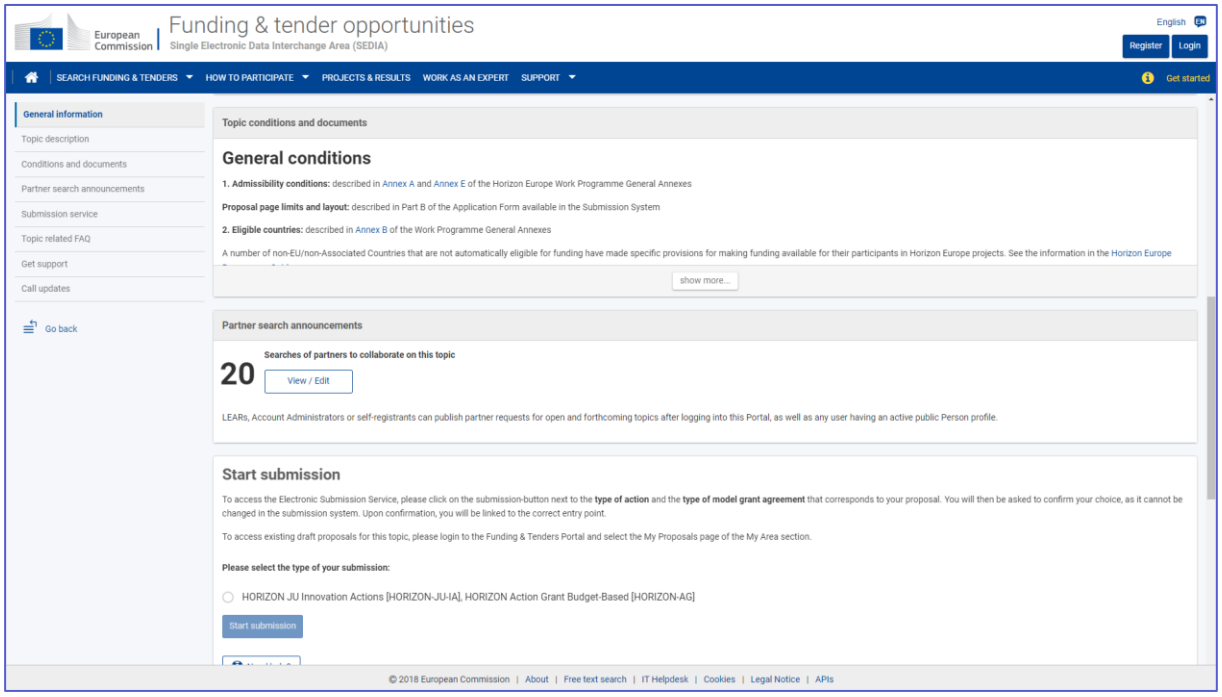

European Commission
Funding & tender opportunities
Single Electronic Data Interchange Area (SEDIA)
English
Register
Login
SEARCH FUNDING & TENDERS
HOW TO PARTICIPATE
PROJECTS & RESULTS
WORK AS AN EXPERT
SUPPORT
Get started
Partner search:
SYNYO GmbH
PIC: 969470478
Austria - WIEN
Small or medium-size enterprise
49 projects funded
Contact organisation
Description
www.synyo.com SYNYO is a reliable partner for your project with profound experience in Horizon Europe. As a SME located in Austria, we are focusing on research, innovation and technology. We explore, develop and implement new methods and solutions with a particular focus on tackling the core challenges of our time. SYNYO supports you to create excellent outcomes in upcoming research and innovation projects by using state-of-the-art approaches, modern technologies and efficient software frameworks. The interdisciplinary team of SYNYO consists of social scientists, engineers, developers and creatives, who are experienced in delivering high-impact results for all phases of RIA, IA and CSA projects. SYNYO uses established tools and self-designed frameworks to implement, communicate, disseminate, and exploit your project in the best way. Additionally, we have proven our management capabilities as the coordinator of several FP7, H2020, HORIZON, CERV and ISF projects, which received excellent evaluations.
Tags/Keywords
. cooperation@synyo.com
. www.synyo.com
AI
AR
Active Assisted Living
Administrative Coordination
Advisory
Ageing
Agency
Analytics
App Development
Application Design
Artificial Intelligence
Augmented Reality
Austria
More tags
Charts pie
Main collaboration partners
Projects list
Partner search announcements
Programmes
REC(1)
ISFP(3)
CERV(2)
H2020(28)
FP7(3)
ISF(2)
HORIZON(9)
ERASMUS2027(1)
Roles
BENEFICIARY(33)
COORDINATOR(16)
© 2018 European Commission | About | Free text search | IT Helpdesk | Cookies | Legal Notice | APIs


European Commission
Funding & tender opportunities
Single Electronic Data Interchange Area (SEDIA)
English
Register
Login
SEARCH FUNDING & TENDERS
HOW TO PARTICIPATE
PROJECTS & RESULTS
WORK AS AN EXPERT
SUPPORT
Get started
Main collaboration partners
14
131
113
84
101
5
2
20
Webtools | © EC-GISCO | Leaflet | © OpenStreetMap contributors | Disclaimer
Projects list
Results: 49
Filtered by: synyo
Search...
TITLE
ACRONYM
PROGRAMME
PROJECT ID
ACTIONS
A RADICAL MODEL OF RESILIENCE FOR YOUNG MINDS
ARMOuR
ISFP
823683
Contact project
Advanced Database for Biomaterials with Data Analysis and Visualisation Tools extended by a Marketplace with Digital Advisors
BIOMATDB
HORIZON
101058779
Contact project
AI-based framework for supporting micro (and small) HSPs on the report and removal of
ALLIES
ISF
101080090
Contact project
© 2018 European Commission | About | Free text search | IT Helpdesk | Cookies | Legal Notice | APIs

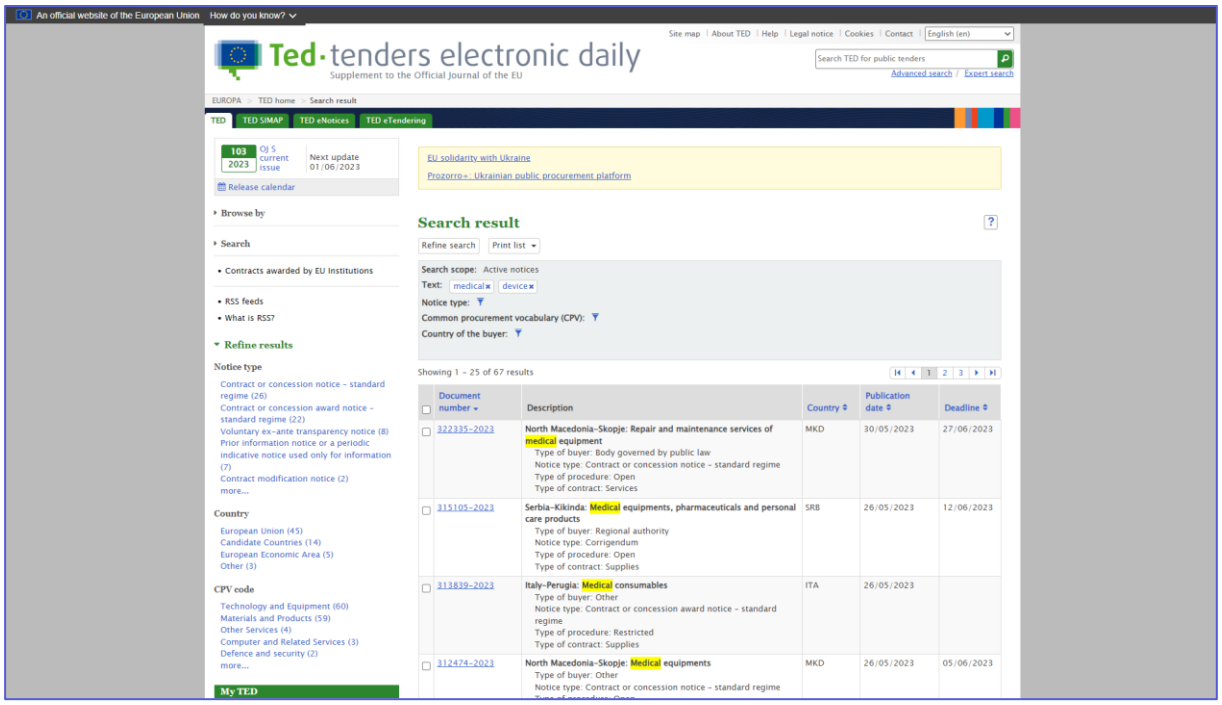
EU Funding: EC Funding & Tender Opportunities

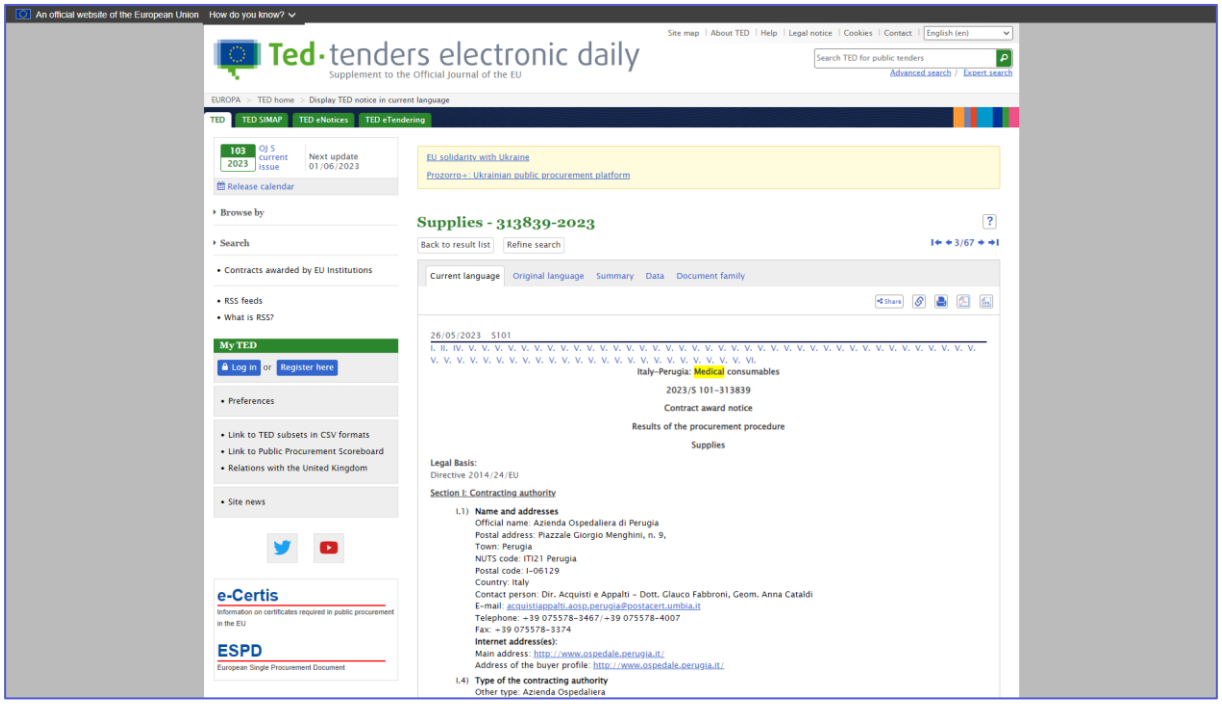





EU Tenders: TED




An official website of the European Union
How do you know?
Ted·tenders electronic daily
Supplement to the Official Journal of the EU
Site map | About TED | Help | Legal notice | Cookies | Contact | English (en)
Search TED for public tenders
Advanced search / Expert search
EUROPA :: TED home :: Display TED notice in current language
TED
TED SIMAP
TED eNotices
TED eTendering
103 2023 OJ S current issue
Next update 01/06/2023
Release calendar
Browse by
Search
Contracts awarded by EU Institutions
RSS feeds
What is RSS?
My TED
Log in or Register here
Preferences
Link to TED subsets in CSV formats
Link to Public Procurement Scoreboard
Relations with the United Kingdom
Site news
e-Certis
Information on certificates required in public procurement in the EU
ESPD
European Single Procurement Document
EU solidarity with Ukraine
Prozorro+: Ukrainian public procurement platform
Supplies - 313839-2023
Back to result list
Refine search
3/67
Current language
Original language
Summary
Data
Document family
Share
26/05/2023 S101
Italy–Perugia: Medical consumables
2023/S 101-313839
Contract award notice
Results of the procurement procedure
Supplies
Legal Basis:
Directive 2014/24/EU
Section I: Contracting authority
I.1) Name and addresses
Official name: Azienda Ospedaliera di Perugia
Postal address: Piazzale Giorgio Menghini, n. 9,
Town: Perugia
NUTS code: ITI21 Perugia
Postal code: I-06129
Country: Italy
Contact person: Dir. Acquisti e Appalti – Dott. Glauco Fabbroni, Geom. Anna Cataldi
E-mail: acquistiappalti.aosp.perugia@postacert.umbria.it
Telephone: +39 075578-3467/+39 075578-4007
Fax: +39 075578-3374
Internet address(es):
Main address: http://www.ospedale.perugia.it/
Address of the buyer profile: http://www.ospedale.perugia.it/
I.4) Type of the contracting authority
Other type: Azienda Ospedaliera