BIOINFORMATICS APPLICATIONS NOTE Vol. 23 no. 5 2007, pages 654–655
doi:10.1093/bioinformatics/btl675

Databases and ontologies

The ESID Online Database network

D. Guzman[1,2], D. Veit[1], V. Knerr[2], G. Kindle[1], B. Gathmann[1], A. M. Eades-Perner[1] and B. Grimbacher[2,*]

[1]Department of Rheumatology and Clinical Immunology, University Hospital Freiburg, Germany and
[2]Department of Clinical Immunology and Molecular Pathology, Royal Free Hospital, University College London, UK

Received on July 19, 2006; revised on December 15, 2006; accepted on January 4, 2007

Advance Access publication January 19, 2007

Associate Editor: Charlie Hodgman

ABSTRACT

Summary: Primary immunodeficiencies (PIDs) belong to the group of rare diseases. The European Society for Immunodeficiencies (ESID), is establishing an innovative European patient and research database network for continuous long-term documentation of patients, in order to improve the diagnosis, classification, prognosis and therapy of PIDs. The ESID Online Database is a web-based system aimed at data storage, data entry, reporting and the import of pre-existing data sources in an enterprise business-to-business integration (B2B). The online database is based on Java 2 Enterprise System (J2EE) with high-standard security features, which comply with data protection laws and the demands of a modern research platform.

Availability: The ESID Online Database is accessible via the official website (http://www.esid.org/).

Contact: b.grimbacher@medsch.ucl.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Primary immunodeficiencies (PIDs) are a group of rare diseases, which usually are the product of genetic defects of the immune system and its development. They represent a key area of study in immunology, providing a better understanding of the immunological pathways, their functions and development. To date, >100 different forms of PIDs have been identified to the molecular level (Shearer and Fischer, 2006). The main approach of identifying the genetic defect, is generating abundant and complex data from different sources, such as knockout mice experiments, genetic diagnosis and case reports. This information is mainly deposited and annotated in peer-reviewed databases, such as OMIM (http://www.ncbi.nlm.nih.gov/omim). While these databases offer a good description of the molecular defect, the correlation with the clinical phenotype is usually poor and limited only to specific case reports. In addition, the clinical phenotype resulting from a given genotype can be quite variable, depending upon many heterogeneous factors. Thus, this relation between genetic and clinical phenotype cannot be assumed to be linear, but is instead a complex expression of molecular defects regulated by endogenous and exogenous factors, which can lead to a complex set of phenotypic features (Buckley, 2005). This overall complexity can be addressed by filling in the existing gap in the clinical phenotype information for PIDs. Given the complexity and low prevalence of PIDs, only through transnational collaborations can sufficient patient numbers be obtained to perform adequate research. The European Society for Immunodeficiencies (ESID), which is a non-profit organization facilitating research on PIDs, has established a European patient and research database network for continuous long-term documentation of patients in order to provide a good platform for scientific research, filling the gap between the generated fundamental knowledge on PIDs and the actual healthcare services delivered to the patients. This secure database network, which is compliant with data protection and international ethics requirements, comprises several novel features and standards for clinical research databases.

The ESID Online Database consists of 206 disease-specific registries. The complete set of registries share a common dataset, which is called 'Core Dataset', handling non-identifying basic patient data, diagnosis, quality of life, therapy, transplantation and laboratory information. In addition to the 'Core Dataset', extended disease-specific datasets are being continuously developed. To date 30 registries include disease-specific extended datasets. The additional information provided by these extended datasets comprise among others aetiology, genetic cause, infection history, additional diagnoses, collected samples, biopsies and surgery events, clinical and additional investigation analyses.

The components of the system used by the ESID Online Database aim at data storage, data entry, reporting and the integration of pre-existing data sources within an enterprise business-to-business integration (B2B) based on Java 2 Enterprise System (J2EE) [Supplementary Material, Fig. 1]. For the data storage component, MySQL MaxDB™(http://www.mysql.com/products/maxdb/) is used. The access to the database is carried out using a Java Database Connectivity interface, JDBC™(http://java.sun.com/products/jdbc/). As J2EE application server, Sun Java™System Application Server (http://www.sun.com/software/products/appsrvr/index.xml) is used. The central component of the system is the Toolwerk Enterprise Integration and Development Platform Application (EIDP) (http://www.toolwerk.de/). EIDP is a multi-tier Java Enterprise Development Platform which provides XML programming interfaces to the different components of a J2EE system, i.e. web tier, EJB tier and services tier. As with the web interface, the EJB tier provides the developer with specific

*To whom correspondence should be addressed.

 © The Author 2007. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

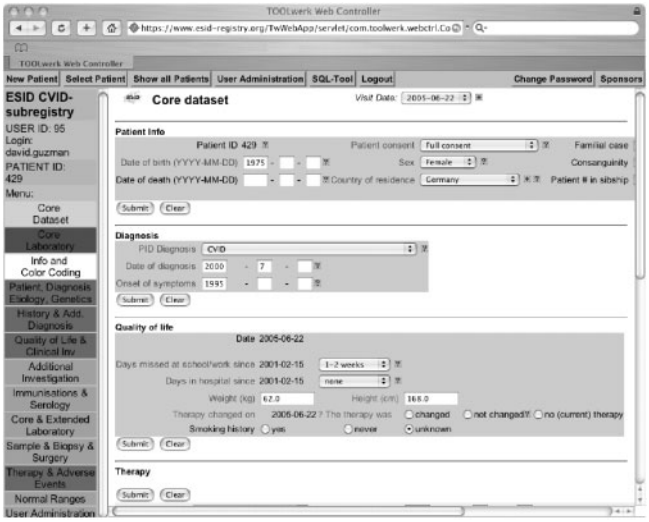

Fig. 1. Screenshot from the ESID Core Dataset user interface, showing the core dataset module. This interface is accessible via a secure login procedure and allows the manual entry of clinical patient data. All available registries are disease specific and access to the patient data is filtered through a central identification number.

functionalities for accessing databases, services and legacy systems. It is even possible to connect different EIDPs together in order to develop complex, secure and high data protection standard systems. The services tier offers specific functionalities for developing the exchange of huge datasets. This tier infrastructure is based on the core of SOAP, but implements its own structures for batch processing. A service builder Java library is available for the development of the clients. The service tier uses the EIDP core system (EJB tier) to access the different data sources [Supplementary Material, Figs 2 and 3].

The ESID Online Database can be accessed using a standard web browser. The access through web browsers is controlled through a user name and password combination. No additional technical prerequisites or plug-ins are required. Communication with other systems is either implemented via SOAP web services calling the EIDP services tier or proprietary communication models when required, where the source database can be accessed through the main EJB tier (plain JDBC, pool connections, other EIDPs or legacy systems) [Supplementary Material, Fig. 4]. All the B2B integration and normal end-user access is only implemented via SSL/TLS encryption. The central server of the ESID Online Database is physically hosted within the secure server network of the IT Centre of the University Hospital Freiburg, Germany.

All the registries within the ESID Online Database are managed through multiple entities [Supplementary Material, Tables 1–7]. The complete set of entities is compliant with the PID list of the International Union of Immunological Societies (Notarangelo *et al*., 2006). With this entity separation, the patients and the data contained are specific to one disease and cannot be accessed through a different application entity. Despite this separation at the application level, the system uses one common database scheme [Supplementary Material, Fig. 5].

The ESID Online Database has an intuitive web interface. A series of web forms are accessible through a side menu containing patient-specific data, which is sorted according to dates where applicable, allowing the documentation of static and follow-up data (Fig. 1).

The ESID Online Database uses a complex and secure permission grant system, which regulates the access to the system resources. Every user is assigned at least a role which controls the display of contents and methods for querying the database. In addition, specific centre and sub-centre codes are also generated, filtering the list of patients available for querying and documenting.

The ESID Online Database also implements advanced logic for more specific tasks, e.g. charts and report generation, search modules, mutation validation, calculations, integration with external databases and elaborated data entry functionalities based on graphical representations. All this logic is implemented through Java code and integrated into the main application web tier as standard plain-old Java objects (POJO), providing a standard mechanism, based on servlets, to combine reusable pieces of Java classes with the session variables and the access to EJB resources [Supplementary Material, Fig. 6].

The use of ICD10 codes (http://www.who.int/classifications/icd/) for disease events descriptions and MeSH terms (available at http://www.nlm.nih.gov/mesh/) for organs and tissues, provides a good platform for the standardized characterization of PIDs, these terms and ontologies are integrated into the web interface, through search functions and forms [Supplementary Material, Fig. 7 and Table 8]. An integration with validated mutation data from the IDbases databases (Riikonen and Vihinen, 1999), is already being implemented, providing cross-linked information to other biological databases, such as Human Genome Database (http://www.gdb.org/gdb/), OMIM, EMBL (http://www.ebi.ac.uk/embl/) and UniProt-Swiss database (http://www.uniprot.org/).

ACKNOWLEDGEMENTS

The ESID Online Database project is sponsored by companies of the Plasma Protein Therapeutics Association Europe (Baxter, Biotest, Grifols, Kedrion, Octapharma) and funded in part by the European Union, EU Contract SP23-CT-2005-006411, EURO-POLICY-PID.

REFERENCES

Buckley,R.H. (2005) Variable phenotypic expression of mutations in genes of the immune system. *J. Clin. Invest*., **115**, 2974–2976.

Notarangelo,L. *et al*. : International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee (2006) Primary immunodeficiency diseases: an update from the International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee Meeting in Budapest, 2005. *J. Allergy Clin. Immunol*., **117**, 883–896.

Riikonen,P. and Vihinen,M. (1999) MUTbase: maintenance and analysis of distributed mutation databases. *Bioinformatics*, **15**, 852–859.

Shearer,W.T. and Fischer,A. (2006) The last 80 years in primary immunodeficiency: how far have we come, how far need we go? *J. Allergy Clin. Immunol*., **117**, 748–752.

SUPPLEMENTARY MATERIALS: The ESID Online Database Network

Guzman D., Veit D., Knerr V., Kindle G., Gathmann B., Eades-Perner AM., Grimbacher B.

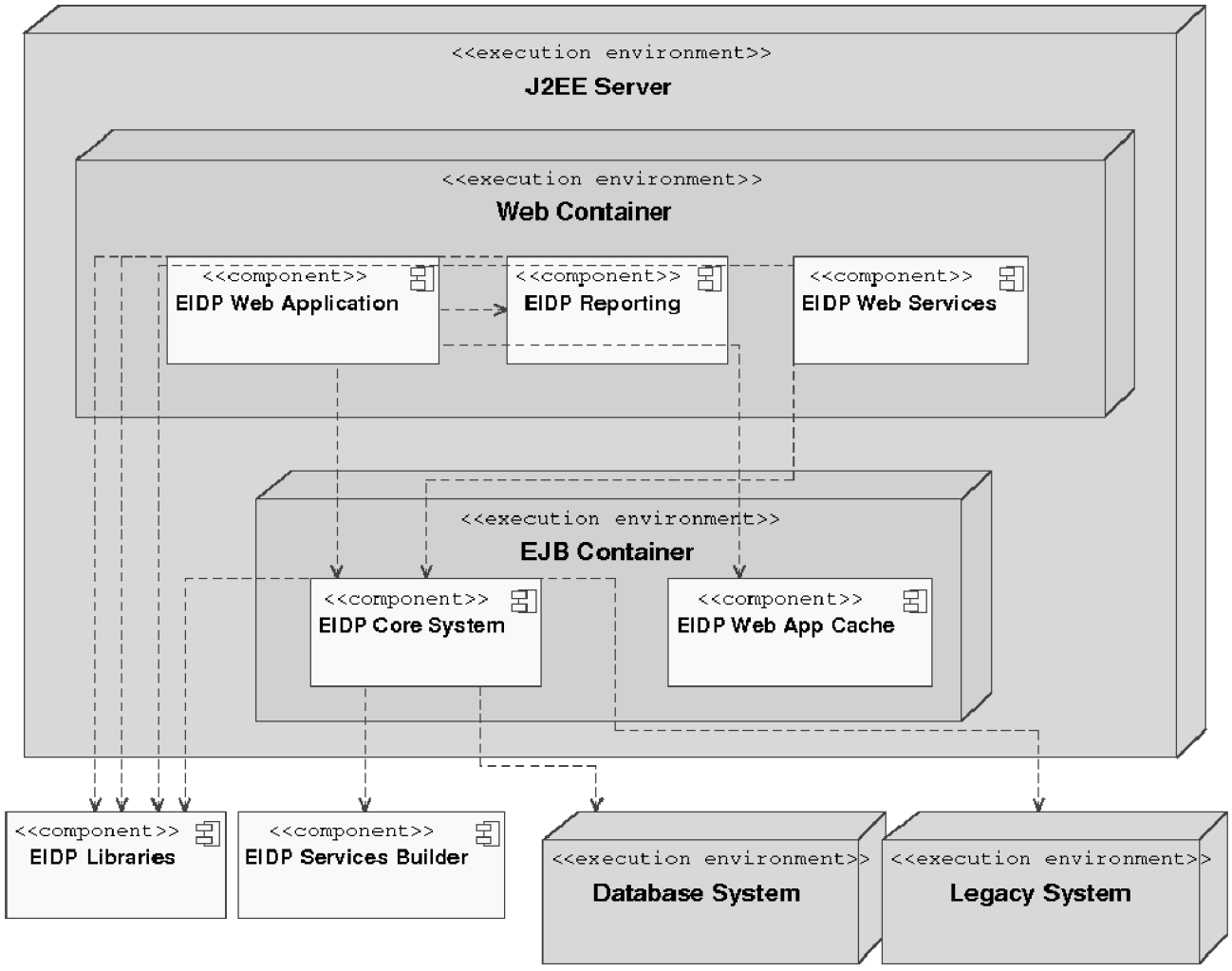

Figure 1: Architecture of the interacting components of the EIDP platform used by ESID Online Database. The Web Container (servlet container) includes the web application, the reporting component and the service infrastructure. All components interact with the Core System, which manages the access to different data sources. The Web App Cache component is a specific infrastructure which stores the data during the application workflow. The components “Web Application”, “Reporting”, “Web Services” and “Core System” are defined through XML programming interfaces, and do not require knowledge of the J2EE framework.

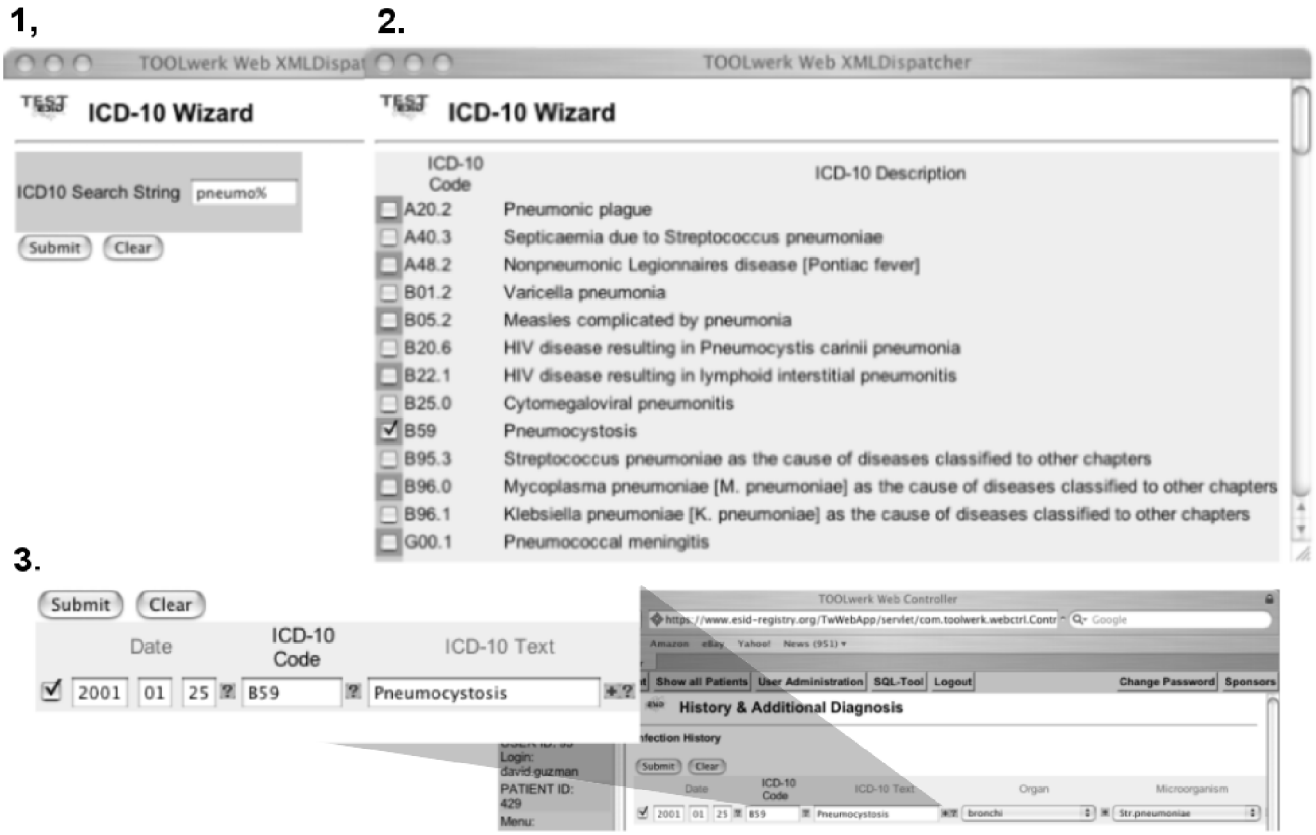

Figure 2: ICD Wizard screenshots from the ESID Online Database. 1. Query interface, allows the users to enter search terms, including the use of SQL wildcards. 2. Search results screen, only one term is allowed to be selected per entry. 3. The selected term is automatically added to the entry field.

Subcategory	*Registry name*
AG	Agammaglobulinemia, X-linked (BTK)
AG	B-cell linker protein deficiency (BLNK / SLP65)
AG	Immunoglobulin heavy Mu chain deficiency (IGHM)
AG	CD79A antigen deficiency
AG	Immunoglobulin Lambda-like polypeptide 1 deficiency (IGLL1)
AG	Agammaglobulinemias with unknown genetic cause
HG	Common variable immunodeficiency (CVID)
HG	Secondary hypogammaglobulinemia
HG	Selective IgM deficiency
HG	Immunoglobulin A deficiency 1, IGAD
HG	Immunoglobulin A deficiency 2, TACI-related
HG	Secondary selective IgA deficiency
HG	Immunoglobulin gene deletions
HG	Isolated IgG subclass deficiency
HG	Immunoglobulin lambda light chain deficiency
HG	Deficiency of specific IgG
HG	Transient hypogammaglobulinemia of infancy
HG	Thymoma with immunodeficiency
HG	ICOS deficiency
HG	CD19 antigen deficiency
HG	BAFF receptor deficiency
HG	Caspase 8 deficiency
HG	Transcobalamin II deficiency
HG	Other Hypogammaglobulinemias
CSR	CD40 antigen deficiency (TNFRSF5)
CSR	CD40 antigen ligand deficiency (CD154)
CSR	Activation-induced cytidine deaminase deficiency (AID)
CSR	Uracil-DNA glycosylase deficiency (UNG)
CSR	Ectodermal dysplasia, anhidrotic, with immune deficiency (IKK-gamma)
CSR	CSR defects and HIGM syndromes with unknown genetic cause

Table 1: Registries in the ESID Online Database contained in the Predominantly Antibody Diseases category. AG=agammaglobulinemias, HG=hypogammaglobulinemias, CSR=class switch recombination defects (hyper IgM syndromes).

Subcategory	*Registry name*
TBM	Recombination-activating gene 1 deficiency (RAG1)
TBM	Recombination-activating gene 2 deficiency (RAG2)
TBM	DNA cross-link repair protein 1C deficiency (Artemis)
TBM	Reticular dysgenesia
TBM	T-B- SCID with unknown genetic cause
TBP	Severe combined immunodeficiency, X-linked (SCIDX1)
TBP	Janus kinase 3 deficiency (JAK3)
TBP	Interleukin 7 receptor deficiency (IL7R)
TBP	Protein-tyrosine phosphatase, receptor-type C deficiency (CD45)
TBP	Protein-tyrosine phosphatase, receptor-type C deficiency (CD45)
TBP	T-B+ SCID with unknown genetic cause
CD8	CD8 antigen alpha polypeptide deficiency (CD8A)
CD4	Selective CD4 cell deficiency
ADA	Adenosine deaminase deficiency (ADA)
PNP	Nucleoside phosphorylase deficiency (NP)
HLA	Major histocompatibility complex class II transactivator (MHC2TA)
HLA	Regulatory factor X, ankyrin-repeat containing (RFXANK)
HLA	Regulatory factor X, 5 (RFX5)
HLA	Regulatory factor X-associated protein (RFXAP)
TAP	ATP-binding cassette transporter 1 deficiency (TAP1)
TAP	ATP-binding cassette transporter 2 deficiency (TAP2)
ZAP	Zeta-chain-associated protein kinase deficiency (ZAP70)
CD3	CD3 antigen gamma subunit deficiency (CD3G)
CD3	CD3 antigen delta subunit deficiency (CD3D)
CD3	CD3 antigen epsilon subunit deficiency (CD3E)
CD3	CD3 deficiency of unknown cause
NU	Winged-helix nude deficiency (FOXN1)
DGS	DiGeorge Syndrome
CMC	Autoimmune polyendocrinopathy syndrome type I (AIRE)
CMC	Other CMC
OB	Other unclassified T-cell disorders

Table 2: Registries in the ESID Online Database contained in the Predominantly T-cell Deficiencies category. TBM=T-B- Severe combined immunodeficiency (SCID), TBP=T-B+ Severe combined immunodeficiency (SCID), CD8=CD8-deficiency, CD4=CD4-deficiency, ADA=ADA-deficiency, PNP=PNP-deficiency, HLA=HLA class II deficiency, TAP=TAP deficiency, ZAP=ZAP deficiency, CD3=CD3-deficiency, NU=NUDE/SCID, DGS=DiGeorge syndrome, CMC=Chronic mucocutaneous candidiasis (CMC), OB=Other unclassified T-cell disorders.

Subcategory	*Registry name*
CGD	Chronic granulomatous disease X-linked (CYBB)
CGD	Chronic granulomatous disease autosomal recessive cytochrome b-negative (CYBA)
CGD	Chronic granulomatous disease autosomal recessive cytochrome b-positive type I (NCF1)
CGD	Chronic granulomatous disease autosomal recessive cytochrome b-positive type II (NCF2)
CGD	CGD with unknown genetic cause
SCN	Elastase 2 deficiency (ELA2)
SCN	Growth factor-independent 1 (GFI1)
SCN	Granulocyte colony-stimulating factor 3 receptor (CSF3R)
SCN	X-linked severe congenital neutropenia (WAS)
CN	Elastase 2 deficiency (ELA2)
SD	SBDS deficiency
AL	Partial albinism and immunodeficiency syndrome
KS	Kostmann syndrome
FLH	Perforin 1 deficiency (PRF1)
FLH	Familial hemophagocytic lymphohistiocytosis 3 (UNC13D)
FLH	Syntaxin 11 deficiency
FLH	Familial hemophagocytic lymphohistiocytosis with unknown genetic cause
GS	Griscelli syndrome type 1 (MYO5A)
GS	Griscelli syndrome type 2 (RAB27A)
GS	Melanophilin deficiency (MLPH)
GS	Griscelli syndrome with unknown genetic cause
CHS	Lysosomal trafficking regulator deficiency (LYST)
CHS	CHS with unknown genetic cause
LAD	Leukocyte adhesion deficiency type 1 (ITGB2)
LAD	Congenital disorder of glycosilation type IIc (FUCT1)
LAD	Leukocyte adhesion deficiency type 3
MPO	Myeloperoxidase deficiency (MPO)
AB	Actin beta deficiency (ACTB)
RAC	RAS-related C3 Bolutinum toxin substrate 2 deficiency (RAC2)
GD	CCAAT/enhancer binding protein epsilon deficiency (CEBPE)
NG6	Glucose-6-phosphate dehydrogenase deficiency (G6PD)
JP	Formyl peptide receptor deficiency
PLS	Cathepsin C deficiency (CTSC)
WHI	Chemokine CXC motif receptor 4 deficiency (CXCR4)
WHI	WHIM syndrome with unknown genetic defect
MYC	Interferon gamma receptor 1 deficiency (IFNGR1)
MYC	Interferon gamma receptor 2 deficiency (IFNGR2)
MYC	Interleukin 12B deficiency (IL12B)
MYC	Interleukin 12 receptor beta-1 deficiency (IL12RB1)

Subcategory	*Registry name*
MYC	Interleukin 18 deficiency (IL18)
MYC	Interleukin 23 alpha deficiency (IL23A)
MYC	Signal transducer and activator of transcription 1 deficiency (STAT1)
MYC	Signal transducer and activator of transcription 5 deficiency (STAT5)
MYC	Susceptibility to mycobacterial infection and unknown genetic defect
OP	Other phagocytic disorders

Table 3: Registries in the ESID Online Database contained in the Phagocytic Disorders category. CGD=Chronic granulomatous disease, SCN=Severe congental neutropenia, CN=Cyclic neutropenia, SD=Schwachman-Diamond-syndrome, AL=PID with partial albinism, KS=Kostmann syndrome, FLH=Familial hemophagocytic lymphohistiocytosis syndromes, GS=Griscelli syndrome, CHS=Chediak Higashi syndrome, LAD=Leukocyte adhesion deficiency, MPO=Myeloperoxidase deficiency, AB=Actin beta deficiency, RAC=RAC2-GTPase defect, GD=Specific granule defect, NG6=Neutrophil glucose-6-phosphate dehydrogenase, JP=Localized juvenile peridontitis, PLS=Papillon-Lefevre syndrome, WHI=Warts hypogammaglobulinemia infections and myelokathexis, MYC=Defects with susceptibility to mycobacterial infection, OP=Other phagocytic disorders.

Subcategory	*Registry name*
CD	Complement component 1, Q subcomponent alpha deficiency (C1QA)
CD	Complement component 1, Q subcomponent gamma deficiency (C1QG)
CD	Complement component C1r deficiency
CD	Complement component 1, s subcomponent deficiency (C1s)
CD	Complement component 2 deficiency
CD	Complement component 3 deficiency (C3)
CD	Complement component 4 deficiency
CD	Complement component 5 deficiency
CD	Complement component 6 deficiency
CD	Complement component 7 deficiency
CD	Complement component 8 deficiency
CD	Complement component 9 deficiency
CD	Properdin P factor complement deficiency (PFC)
CD	Complement factor B deficiency
CD	Factor D deficiency
CD	I Factor deficiency (IF)
CD	Complement factor H deficiency
CD	Hereditary Angioedema (C1inh)
CD	C3b inactivator deficiency
CD	Mannan-binding lectin serine protease 2 deficiency (MASP2)
CD	Decay-accelerating factor for complement deficiency (DAF CD55)
CD	CD59 antigen P18-20 deficiency (CD59)
MBL	Mannose-binding lectin deficiency

Table 4: Registries in the ESID Online Database contained in the Complement Deficiencies category. CD=Complement deficiency, MBL=Mannose-binding lectin.

Subcategory	*Registry name*
XLP	Lymphoproliferative syndrome, X-linked (SH2D1A)
XLP	XLP with unknown genetic cause
IGE	Hyper-IgE syndrome
DNA	Ataxia telangiectasia (ATM)
DNA	AT-like disorder
DNA	Nijmegen breakage syndrome (NBS1)
DNA	Bloom syndrome (RECQ2)
DNA	Fanconi anemia
DNA	Seckel syndrome
DNA	DNA-ligase 4, ATP-dependent deficiency (LIG4)
DNA	Other DNA-breakage disorder
WAS	Wiskott-Aldrich syndrome with mutations in WASP
WAS	Wiskott-Aldrich syndrome with unknown genetic cause
OST	Osteopetrosis, chloride channel 7 deficiency (CLCN7)
OST	Osteopetrosis-associated transmembrane protein 1 deficiency (OSTM1)
OST	Osteopetrosis with renal tubular acidosis (CA2)
OST	Osteopetrosis, T-cell immune regulator 1 deficiency (TCIRG1)
ICF	DNA methyl transferase 3B deficiency (DNMT3B)
ICF	ICF with unknown genetic cause
CH	Mitochondrial RNA-processing endoribonuclease deficiency (RMRP)
SD	Immunoosseus dysplasia, Schimke type (SMARCAL1)
HP	Adaptor-related protein complex 3 beta 1 subunit deficiency (AP3B1)
FC	Fc fragment for IgG high affinity receptor Ia defect (FCGR1A CD64)
FC	Fc fragment for IgG low affinity receptor IIa defect (FCGR2A CD32)
FC	Fc fragment for IgG low affinity receptor IIb defect (FCGR2B CD32)
FC	Fc fragment for IgG low affinity receptor IIIa defect (FCGR3A CD32)
FC	Fc fragment for IgG low affinity receptor IIIb defect (FCGR3B CD32)
FC	Fc fragment for IgG receptor transporter alpha defect (FCGRT)
NS	Serine protease inhibitor Kazal-type 5 deficiency (SPINK5)
NS	Netherton syndrome with unknown genetic cause
KB	Ectodermal dysplasia, anhidrotic, with immune deficiency (IKK-gamma)
KB	Nuclear factor of kappa light chain gene enhancer in B cells inhibitor alpha defect (NFKBIA)
KB	Interleukin 1 receptor-associated kinase 4 deficiency (IRAK4)
EV	Epidermodysplasia verruciformis gene 1 deficiency (EVER1)
EV	Epidermodysplasia verruciformis gene 2 deficiency (EVER2)

Table 5: Registries in the ESID Online Database contained in the Other Well Defined PIDs category. XLP=X-linked lymphoproliferative syndrome, IGE=Hyper IgE syndrome, DNA=DNA-breakage disorder, WAS=Wiskott-Aldrich syndrome, OST=Osteopetrosis, ICF=Immunodeficiency centromeric instability facial anomalies syndrome, CH=Cartilage hair hypoplasia, SD=Schimke disease, HP=Hermansky-Pudlak syndrome 2, FC=Fc receptor deficiencies, NS=Netherton syndrome, KB=Defects of TLR/NFkappa-B signalling, EV=Epidermodysplasia verruciformis.

Subcategory	*Registry name*
ALP	Tumor necrosis factor receptor superfamily member 6 deficiency (ALPS IA-type)
ALP	Tumor necrosis factor ligand superfamily member 6 deficiency (ALPS IB-type)
ALP	Caspase 8 deficiency (ALPS IIB-type)
ALP	Caspase 10 apoptosis-related cysteine protease deficiency (ALPS IIA-type)
ALP	ALPS with unknown genetic cause
APE	Autoimmune regulator deficiency (AIRE)
APE	APECED with unknown genetic cause
IPX	Foxkhead box P3 deficiency (FOXP3)
IPX	IPEX with unknown genetic cause
MF	Familial mediterranean fever defect (MEFV)
PF	Familial periodic fever, autosomal dominant (TRAPS)
PF	Hyper IgD syndrome (MVK)
MW	CIAS1 gene defect
FC	CIAS1 gene defect
CS	CIAS1 gene defect
PYO	Proline/serine/threonine phosphatase-interacting protein 1 deficiency (PSTPIP1)
BLA	Caspase recruitment domain-containing protein 15 deficiency (CARD15)

Table 6: Registries in the ESID Online Database contained in the Autoimmune and immunedysregulation syndromes category. ALP=Autoimmune lymphoproliferative syndrome (ALPS), APE=Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED), IPX=Immunodysregulation polyendocrinopathy enteropathy X-linked (IPEX), MF=Familial mediterranean fever, PF=Familial periodic fever, MW=Muckle-Wells syndrome, FC=Familial cold autoinflammatory syndrome, CS=CINCA syndrome, PYO=Pyogenic sterile arthritis pyoderma gangrenosum and acne, BLA=Blau syndrome.

ICD10 Code	*No. Cases*	*ICD10 Term*
J40	152	Bronchitis, not specified as acute or chronic
J32	123	Chronic sinusitis
J15	63	Bacterial pneumonia, not elsewhere classified
J00	38	Acute nasopharyngitis [common cold]
H66	26	Suppurative and unspecified otitis media
J47	23	Bronchiectasis
H10	21	Conjunctivitis

Table 7: Representative ICD10 terms, describing the infection events of patients affected by CVID in the ESID Online Database. These values are not mutually exclusive, more than one infection event can be documented for a given patient.