



PORTABERTA

Portal de Investigação para a Ciência Aberta

Atividade 2 - Estudos técnicos, de diagnóstico e prospetivos

E2.1 - Especificação dos requisitos do sistema e gestão de entidades

Versão: 1.0

Autor(es): Raquel Truta, Pedro Príncipe, Ricardo Saraiva

Revisão: Eloy Rodrigues

Data: setembro de 2021

Documento público

PortAberta - Portal de Investigação para a Ciência Aberta | Operação 000431 (SAMA2020).

Cofinanciado pelo Programa Operacional Competitividade e Internacionalização (COMPETE 2020) do Portugal 2020 e pela União Europeia através do Fundo Social Europeu.

Índice

INTRODUÇÃO	3
VISÃO GERAL DO VIVO	3
PRINCIPAIS FUNÇÕES	4
<i>O que permite o VIVO</i>	4
ESPECIFICAÇÃO DOS REQUISITOS DO SISTEMA	5
ESPECIFICAÇÕES	5
<i>Requisitos do sistema</i>	5
<i>Recomendações de Hardware</i>	5
<i>Recomendações de Software</i>	5
<i>Arquitetura</i>	6
GESTÃO DE ENTIDADES	6
COMMON EUROPEAN RESEARCH INFORMATION FORMAT (CERIF)	6
<i>Tipos de entidades conceptuais e características CERIF</i>	7
VIVO	8
<i>Modelo de dados</i>	8
FONTES DE INFORMAÇÃO PARA INTEGRAÇÃO E GESTÃO DE ENTIDADES	9
<i>Entidades vs Fontes de informação</i>	9
<i>Esquemas dos modelos de dados no VIVO para ilustrar a parametrização do sistema</i>	11

Introdução

O portal de investigação para a Ciência Aberta – Porta Aberta, entre outros aspetos, visa permitir que os utilizadores giram a sua produção científica através do registo e/ou importação de informação sobre as suas publicações, comunicações e outros resultados científicos e académicos.

Cada autor terá o seu próprio perfil público onde pode gerir a informação pública como os contactos, as ligações externas, a informação curricular, as qualificações académicas, as atividades de ensino, as orientações, a produção científica e redes de colaborações.

O portal também dará acesso a estatísticas de produção científica do conjunto das instituições participantes e das suas unidades internas, tais como as unidades de investigação, as escolas e os departamentos. Cada autor poderá comparar o seu trabalho com o trabalho do investigador médio em diferentes unidades da instituição, tendo também acesso às citações e dados sobre as revistas das suas publicações.

O VIVO¹ É um software de código aberto, apoiado pela Comunidade Lyris, para representar a atividade académica. O VIVO apoia a gravação, edição, pesquisa, navegação, e visualização da atividade académica. Encoraja o registo académico, a investigação e a avaliação do impacto da mesma. O VIVO é facilmente adaptado para apoiar domínios adicionais da atividade académica.

O software Duraspace VIVO é uma aplicação web semântica *open source*, inicialmente desenvolvida pela Universidade Cornell nos Estados Unidos da América e que, quando instalada e populada com os interesses, atividades e concretizações dos investigadores, permite a descoberta de investigação e conhecimento transversal às disciplinas nessa instituição e não só.

O VIVO é um sistema de gestão de informação académica.

Exemplo a considerar de instalações:

- <https://scholars.library.tamu.edu/vivo/>
- <https://vivo.tib.eu/fis/>
- <https://scholars.duke.edu/>

VIVO - connect | share | discover - <https://duraspace.org/vivo/>

Visão geral do VIVO

O conteúdo de uma instalação VIVO pode ser mantido manualmente, ou de forma automatizada a partir de sistemas locais de registo, tais como RH, bolsas, bases de dados de cursos e atividades dos docentes, ou de fornecedores de bases de dados, tais como agregadores de publicações e agências de financiamento.

¹ VIVO: <https://duraspace.org/vivo/>

Existem diversas atividades associadas ao projeto VIVO, através de agências federais, instituições académicas, sociedades profissionais, editoras com fins lucrativos, e fornecedores de dados, bem como uma variedade de esforços com as comunidades de desenvolvimento da web semântica e de ontologia.

Parceiros de maior relevo: [CASRAI](#) (Consortium Advancing Standards in Research Administration Information), [EuroCRIS](#) (Current Research Information Systems) e a Iniciativa [ORCID](#) (Open Researcher and Contributor ID).

O VIVO é uma rede em que mais de 140 instituições e agências estão presentes em mais de 25 países.

- **Open Source | Código aberto**

O VIVO e todos os componentes da VIVO são fornecidos como fonte aberta. Download da informação a partir do GitHub;

- **Open Community | Comunidade aberta**

A Comunidade VIVO é aberta para todos. É possível seguir os trabalhos do VIVO na sua Wiki;

- **Open Data | Dados abertos**

O VIVO produz dados abertos interligados que podem ser facilmente partilhados e combinados.

Principais funções

- Acesso online
 - o VIVO disponibiliza um portal online onde é possível consultar a Comunidade Académica, o seu trabalho e as suas interligações.
- Obtenção de dados
 - disponibilização de determinados dados conforme as permissões atribuídas aos diferentes intervenientes
- Controlo de acesso
 - possibilidade de autenticação através de Shibboleth ou com base em palavra-chave

Informações detalhadas em:

<https://wiki.lyrasis.org/display/VIVODOC112x/VIVO+1.12.x+Documentation>

<https://wiki.lyrasis.org/display/VIVODOC111x/Installing+VIVO#InstallingVIVO-Overview>

O que permite o VIVO

- todos os componentes de fonte aberta e a instituição possuir/controlar os seus dados
- solução de fácil instalação e/ou alojamento (software como um serviço)
- perfis simplificados que recolhem dados de um número limitado de fontes
- interface de utilizador otimizada para uma maior acessibilidade
- editor de perfis melhorado
- apoiar a reutilização e/ou comunicação de dados

Especificação dos requisitos do sistema

O VIVO é um sistema de gestão de informação académica. Recolhe e armazena dados estruturados sobre a investigação do corpo docente e as atividades académicas, de modo a que a informação possa ser apresentada de várias formas:

- **Autoridades** - ingerir dados de fontes autorizadas;
- **Normas** - para representar a atividade académica e permitir a interoperabilidade, a importação de dados externos e a reutilização dos dados do sistema;
- **Casos de uso** - apoiam a visualização da história académica, disponibiliza a investigação, permite pesquisa de peritos, análise de rede e avaliação do impacto da investigação.

Especificações

Requisitos do sistema

- Recomendações de Hardware
- Requisitos do software

Para mais especificidades técnicas consultar:

<https://wiki.lyrasis.org/display/VIVODOC112x/System+Requirements>

Recomendações de Hardware

Na configuração do VIVO é recomendada uma largura de banda alta para o sistema de arquivos usado pelo MySQL e memória significativa para camadas de cache do banco de dados.

- Especificações mínimas: 2 core x64 processador, 2GB RAM, 100GB HDD
- Especificações recomendadas: 4 core x64 processador, 16GB RAM, 500GB SSD

ATENÇÃO: “I/O performance for MySQL is critical to the responsiveness of the application. The fastest SSD you can specify will help, as will having direct (e.g. not virtualised) access to it”.

Recomendações de Software

Sistemas Operativo:

VIVO é agnóstico do Sistema operativo (SO), pode por isso ser executado em qualquer SO, no entanto é sugerido LINUX e em função de ser um aplicativo Java, depende de ter uma Java Virtual Machine e também o Tomcat servlet container.

Para glibc 2.14+ (Linux only) a documentação sugere:

Distribution	Ubuntu	Debian	Fedora	CentOS	RHEL	SUSE	Gentoo
Minimum version	12	8	16	7	7	12	12

Componentes requeridos:

- Java 8 or 11 (*the minimum requirement is Java 8*).
- Maven 3.6.1 or later (*the installation mechanism uses Maven to package and deploy the VIVO application and other necessary files - the minimum version of Maven required is 3.6.1, although it is better to use a more recent version of the 3.x releases where possible*).
- MySQL / MariaDB 5.5 or later (or any other supported by Jena SDB).
- Tomcat 8.5 or later.

Arquitetura

O VIVO é um sistema de software que depende de numerosos componentes de software de código aberto. Fundamentalmente, a VIVO baseia-se no Vitro². VIVO adiciona uma [coleção de ontologias](#) para representar dados sobre o sistema académico.

O Vitro é um motor de web semântica de código aberto, de uso geral. É a plataforma de desenvolvimento de aplicações subjacente ao VIVO. O Vitro não tem conhecimento de domínio. Dadas as ontologias relativas a um domínio, o Vitro apoia a edição da ontologia, criação de indivíduos, gestão de indivíduos em "páginas" que gera, organização de indivíduos em "grupos de classe", indexação, pesquisa, navegação facetada, consulta, importação e exportação.

O VIVO é Vitro com uma ontologia para representar bolsas de estudo, e um conjunto de exibições e visualizações que apoiam a utilização de dados para a descoberta de especialistas, formação de equipas, avaliação, e outros casos de utilização da VIVO.

O Vitro pode ser construído e operado independentemente do VIVO, no entanto, o VIVO está completamente dependente do Vitro.

Mais informação em: <https://wiki.lyrasis.org/display/VIVODOC19x/Architecture>.

Gestão de entidades

Common European Research Information Format (CERIF)

A CERIF é uma norma recomendada pela União Europeia aos seus Estados Membros³. O modelo físico da CERIF é um modelo de base de dados relacional disponível como scripts SQL baseados em construções ERM (Entity Relationship Model) comuns. Os últimos lançamentos incluem uma "Camada Semântica" formalizada e um formato de intercâmbio XML.

² Vitro: <https://wiki.lyrasis.org/display/VIVODOC111x/Vitro>

³ CERIF: <https://cordis.europa.eu/article/id/8260-cerif-common-european-research-information-format>



Figura 1. Entidades CERIF e as suas relações

Tipos de entidades conceptuais e características CERIF

O modelo CERIF está conceptualmente estruturado em tipos e características de entidades. Entre os tipos, distinguimos entidades **nucleares (core)**, de **resultado (result)**, de **ligação (link)**, e de **segundo nível (2nd)**. Como características, consideramos o multilinguismo e a semântica.

- **Entidades CERIF de núcleo (core):** As entidades de núcleo são *'Person'*, *'OrganizationUnit'* e *'Project'*. Elas permitem a representação de atores científicos. A figura 1 mostra-as no centro inferior, indicando as suas relações recursivas (círculos) e de ligação. Cada entidade do núcleo liga-se a si própria e mantém relações com muitas outras entidades.
- **Entidades CERIF de Resultado (result):** As entidades de resultado são *'ResultPublication'*, *'ResultPatent'* e *'ResultProduct'*. Elas permitem a representação dos resultados da investigação. A figura 1 mostra-as no centro superior, indicando as suas relações.
- **Entidades CERIF de 2º nível (2nd level):** As entidades de 2º nível são, *'Funding'*, *'Facility'*, *'Equipment'*, *'Prize'*, *'CV'*, *'Expertise'*, *'Qualification'*, *'Citation'*, *'Metrics'*, *'Event'*, *'PostalAddress'*, and *'ElectronicAddress'*. Elas permitem a representação do ambiente de investigação. A figura 1 mostra as entidades de segundo nível que rodeiam o núcleo e as entidades de resultados.
- **Entidades de ligação CERIF (link):** As entidades de ligação são consideradas uma das principais forças do modelo CERIF. Entidades de ligação são as relações concretizadas entre entidades do núcleo, do resultado, e do 2º nível. Uma entidade de ligação liga sempre duas entidades e inclui uma referência com carimbo de data/hora a um classificador que é ela própria atribuído a um esquema de classificação.
- **Características multilinguísticas CERIF (lang):** A CERIF suporta múltiplas características linguísticas para nomes, títulos, descrições, palavras-chave, resumos, e mesmo para a semântica.

- **Características semânticas CERIF (classe):** A camada semântica CERIF é considerada um recipiente para gerir e manter a semântica formal (contextos), tal como estabelecido com as entidades de ligação. Permite a representação de tipos de relação, vistas de aplicação, títulos de assunto, qualquer esquema de classificação, ou mesmo o mapeamento entre esquemas.

A estrutura conceptual dos tipos e características do CERIF é apenas uma estrutura virtual e, como tal, não inerente ao modelo de dados físicos. Destina-se a apoiar a compreensão do modelo.

VIVO

Modelo de dados

O "modelo de dados" é frequentemente referido no VIVO, mas poderá ser vista como uma simplificação excessiva. De facto, o VIVO contém uma matriz de modelos e submodelos de dados, gráficos, conjuntos de dados e outras construções.

Será mais correto falar sobre a união destes modelos de dados como a 'base de conhecimento' (*knowledge base*). Contudo, a terminologia de "modelo de dados" está firmemente enraizada no VIVO.

A partir da versão 1.6 do VIVO, começou-se a tentar simplificar esta complexa coleção de modelos, e a produzir uma camada de acesso unificado. Sendo, em curso, independentemente de quão limpo o desenho possa eventualmente tornar-se, esta continuará a ser uma área com requisitos complexos que não podem ser satisfeitos por soluções simplistas.

Mais Informação em: <https://wiki.lyrasis.org/display/VIVODOC111x/VIVO+Data+Models>

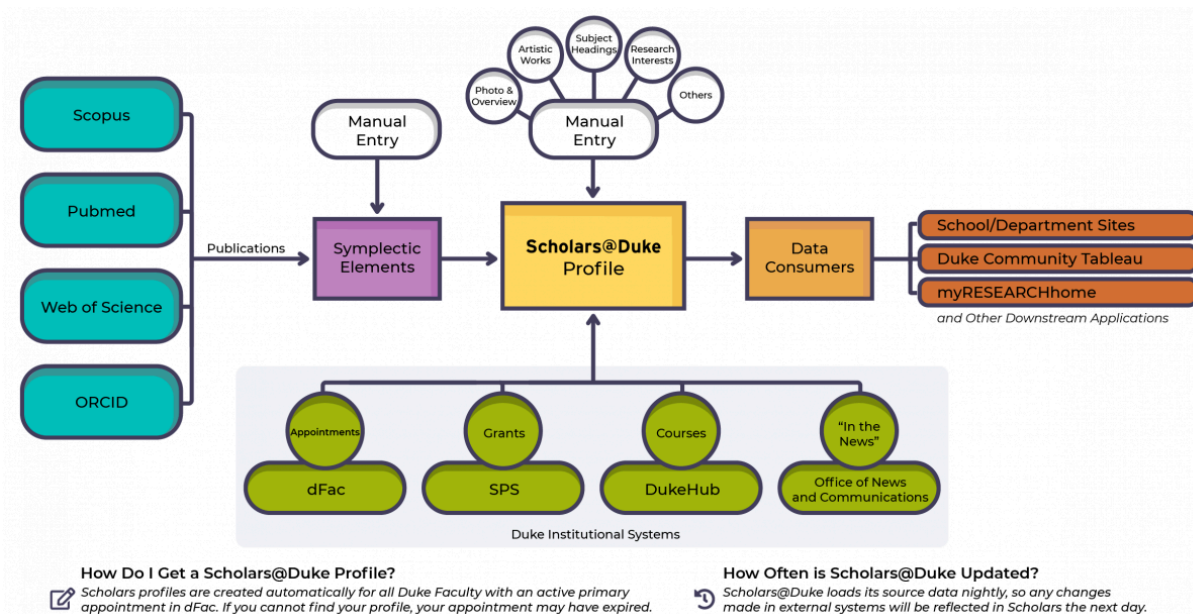


Figura 2. Exemplo de implementação VIVO na Duke University⁴

⁴ Scholars@Duke: <https://scholars.duke.edu/>

Fontes de informação para integração e Gestão de entidades

Entidades vs Fontes de informação

Tabela ilustrativa tendo por base tipos de entidades conceptuais e características CERIF

Entidade(s)	Tipo	Fonte(s) de Informação	PIDs / APIs	OBS
Pessoas	Núcleo (core) <i>'Person'</i>	Interna: Institucional (RH) Externa: ORCID ID CIENCIA ID	Nº Mecanográfico ⁵ Login ⁶ ORCID ID CIENCIA ID Researcher ID Scopus ID Google ID	UMinho - N. Mec IPB - Login
Unidades de investigação	Núcleo (core) <i>'OrganizationUnit'</i>	Interna: Institucional (RH)	FCT/FCCN Ringgold ISNI ROR	Universidade Faculdade Departamento Centro de Investigação
Projetos	Núcleo (core) <i>'Project'</i>	Interna: Institucional (Projetos de investigação) Externa: Financiadores de Ciência	OpenAIRE API Sciproj FCT	Nacionais Internacionais
Resultados de investigação	Resultado (result) <i>'ResultPublication'</i>	Interna: Repositórios de publicações e de dados, Publicações Académicas ⁷ Externa: CVITAE ORCID	Handle DOI	Publicações Patentes Dados Software Etc..
	Resultado (result) <i>'ResultProduct'</i>			
	Resultado (result) <i>'ResultPatent'</i>	PATENTSCOPE ⁸		

⁵ A Universidade do Minho utiliza números mecanográficos para identificação/credenciação (estão organizados por tipologias ex. d0000 = docente, b0000=bolseiro)

⁶ Login IPB

⁷ Universidade do Minho

⁸ <https://patentscope.wipo.int/search/en/search.jsf>

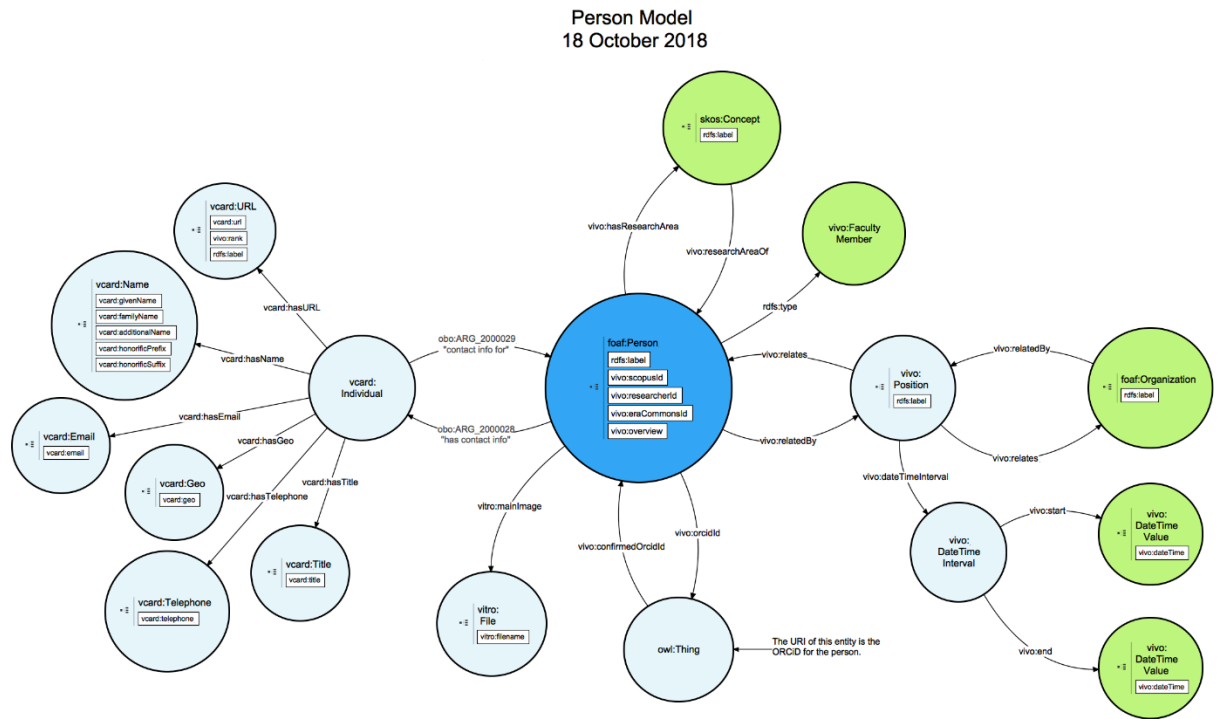
Financiamento	2º nível (2nd level) <i>'Funding'</i>	Interna: Institucional	Mod. "Gestão Projetos"/Portal UMinho ⁹ Projetos de Investigação ¹⁰ (Unidade(s) Financeira, apoio a projetos)	
Instalações	2º nível (2nd level) <i>'Facility'</i> ,	Interna: Institucional		
Equipamento	2º nível (2nd level) <i>'Equipment'</i>	Interna: Institucional		
Prémios	2º nível (2nd level) <i>'Prize'</i>	Individual (edição do perfil no portal) Interna: Institucional (Centros de investigação, Investigadores, Gabinete de Comunicação e Imagem)		Election to learned society; Fellowship awarded competitively; National/international honour; Other distinction; Prize (including medals and awards)
CV	2º nível (2nd level) <i>'CV'</i>	Individual (edição do perfil no portal) Externa: Ciência Vitae ORCID		
Especialidade(s)	2º nível (2nd level) <i>'Expertise'</i>	Individual (edição do perfil no portal) Externa: Ciência Vitae ORCID		
Qualificações	2º nível (2nd level) <i>'Qualification'</i>	Individual (edição do perfil no portal) Interna: Institucional (RH) Externa: Ciência Vitae		

⁹ Universidade do Minho

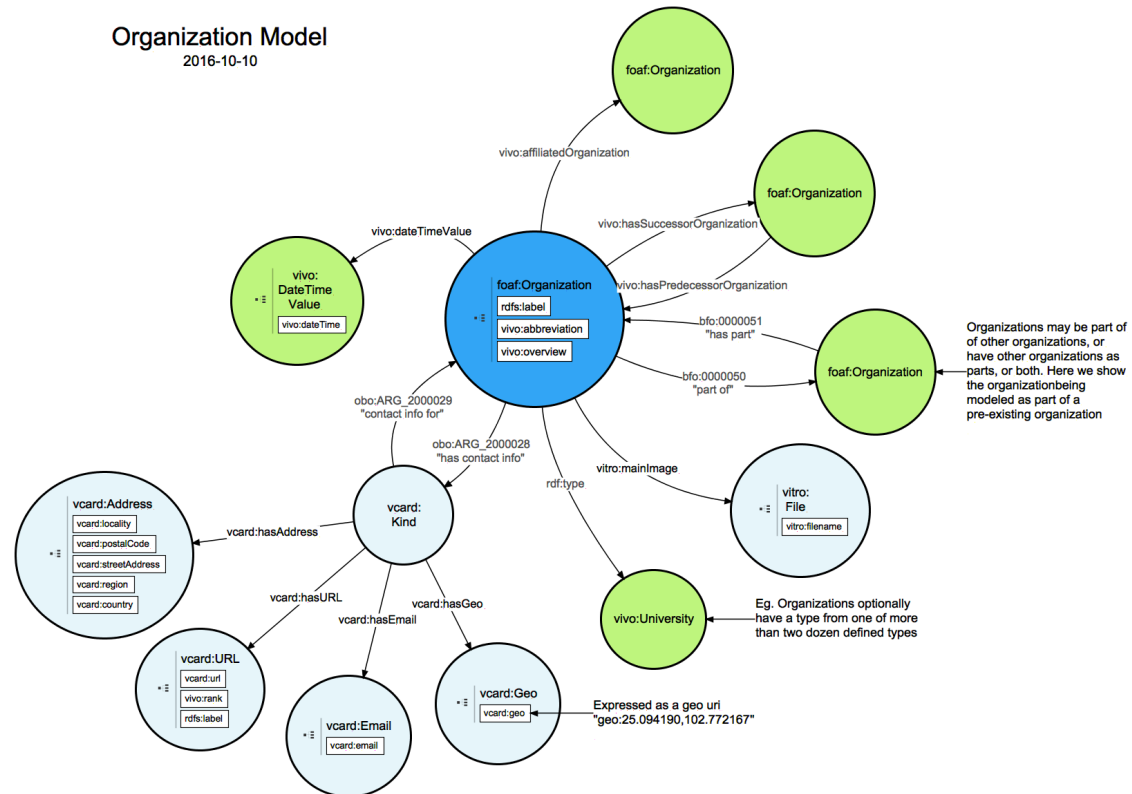
¹⁰ IPB

Citações	2º nível (2nd level) <i>'Citation'</i>	Externa: WoS SCOPUS PubMed Open Citations GScholar		
Métricas	2º nível (2nd level) <i>'Metrics'</i>	Externa: Dimensions Altmetrics		
Endereço Postal	2º nível (2nd level) <i>'PostalAddress'</i> ,	Interna: Institucional (RH)		
Endereço Electrónico	2º nível (2nd level) <i>'ElectronicAddress'</i>	Interna: Institucional (RH)		
País	2º nível (2nd level) <i>'Country'</i>	Interna: Institucional (RH)		
Outras a considerar não incluídas no modelo de dados VIVO				
Evento	2º nível (2nd level) <i>'Event'</i> ,	Individual (edição do perfil no portal) Interna: Institucional (Centros de investigação, Investigadores, Gabinete de Comunicação e Imagem)		Entidade parte do modelo CERIF, a considerar no VIVO
Atividades		Individual (edição do perfil no portal)		Exemplo de algumas previstas no PURE e a considerar no VIVO: Consultancy Editorial activity Membership of board Membership of committee Membership of council
Notícias/Media		Interna: Institucional (Centros de investigação, Investigadores, Gabinete de Comunicação e Imagem)		Título Âmbito Media Tipo de media País Data URL

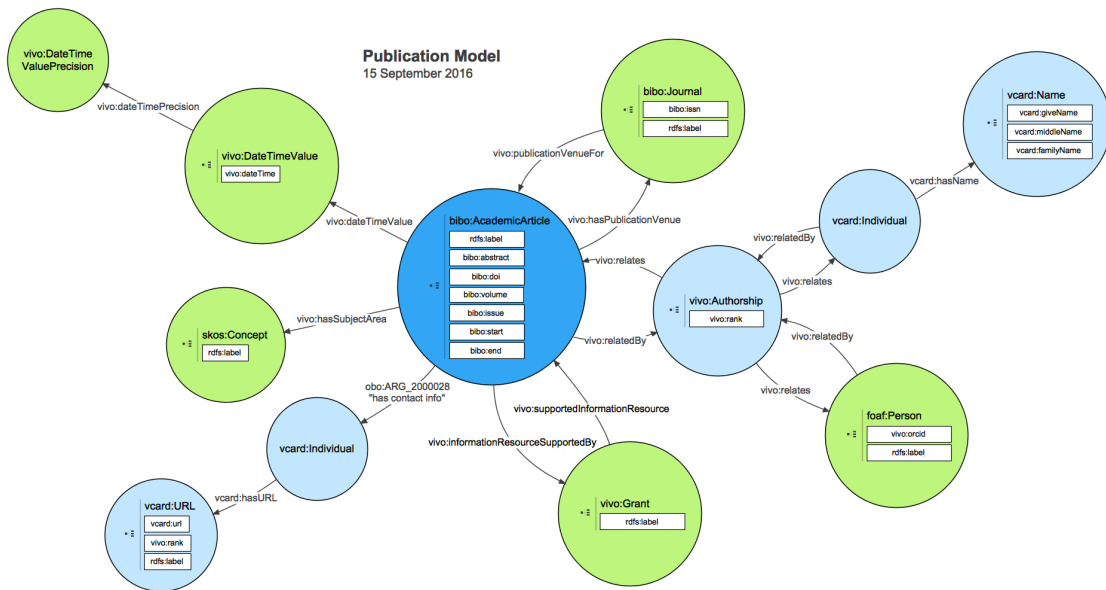
Esquemas dos modelos de dados no VIVO para ilustrar a parametrização do sistema



<https://wiki.lyrasis.org/display/VIVODOC111x/Person+Model>

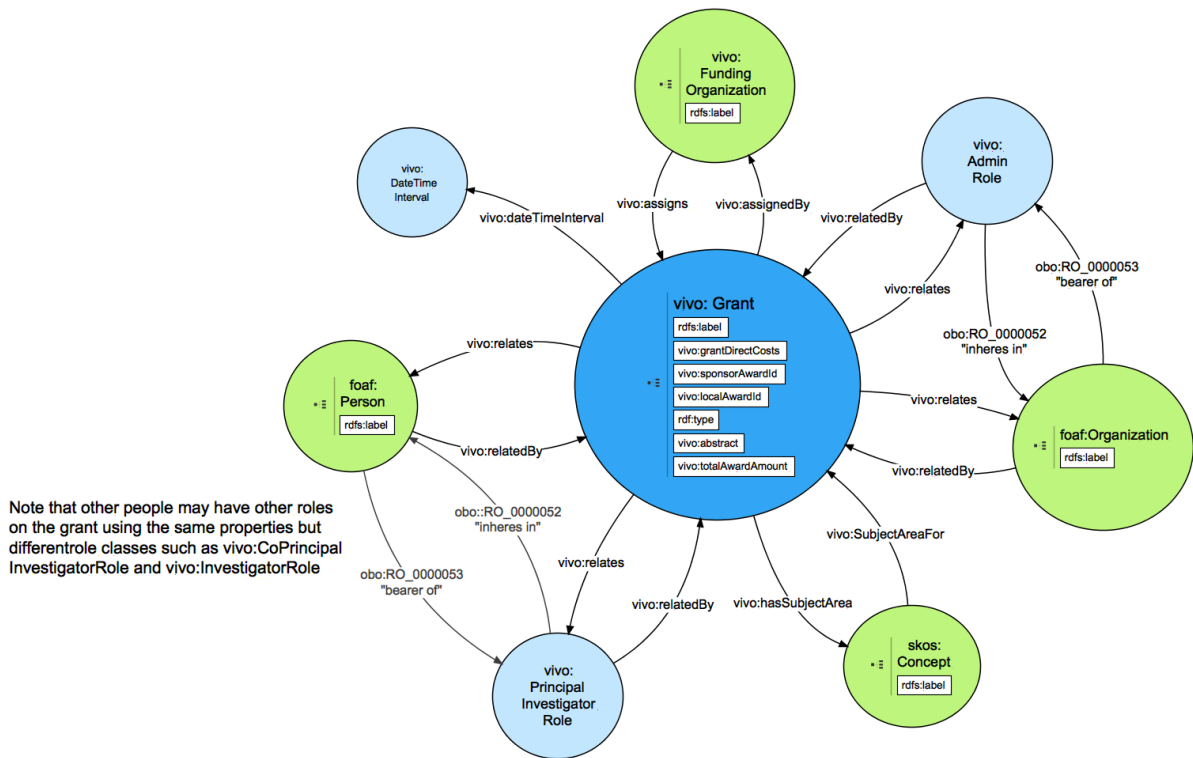


<https://wiki.lyrasis.org/display/VIVODOC111x/Organization+Model>



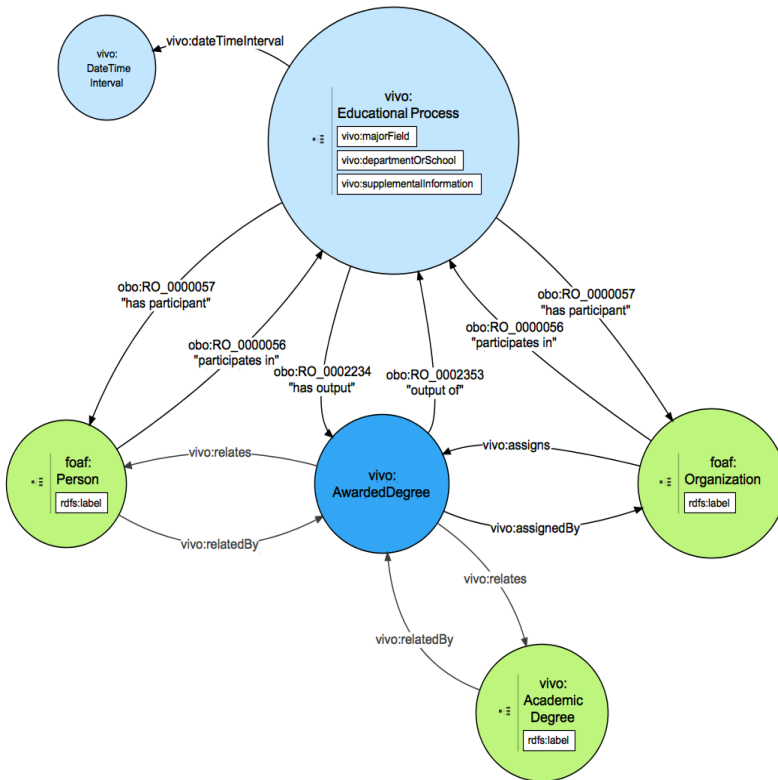
<https://wiki.lyrasis.org/display/VIVODOC111x/Publication+Model>

Grant Model
12-October-2016

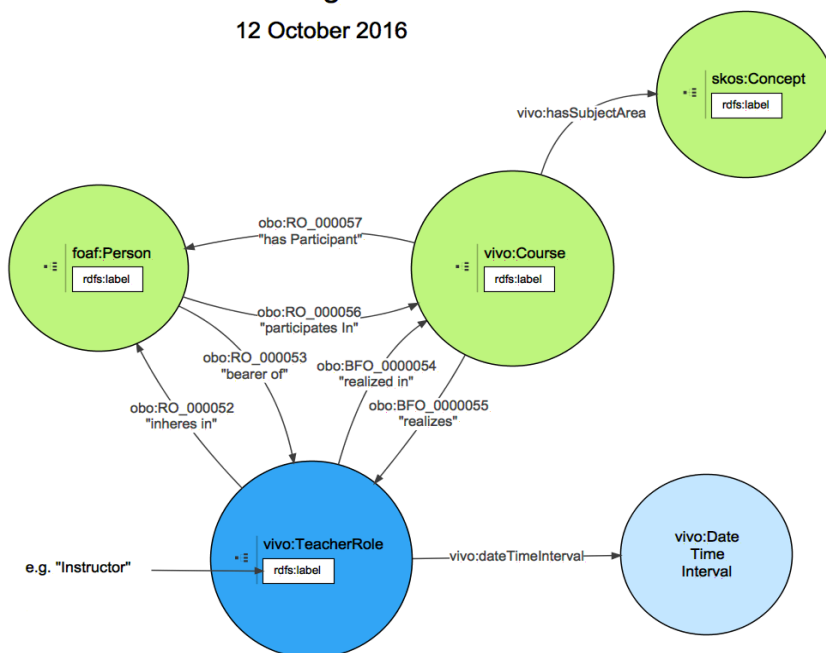


<https://wiki.lyrasis.org/display/VIVODOC111x/Grant+Model>

Education and Training Model 12 October 2016

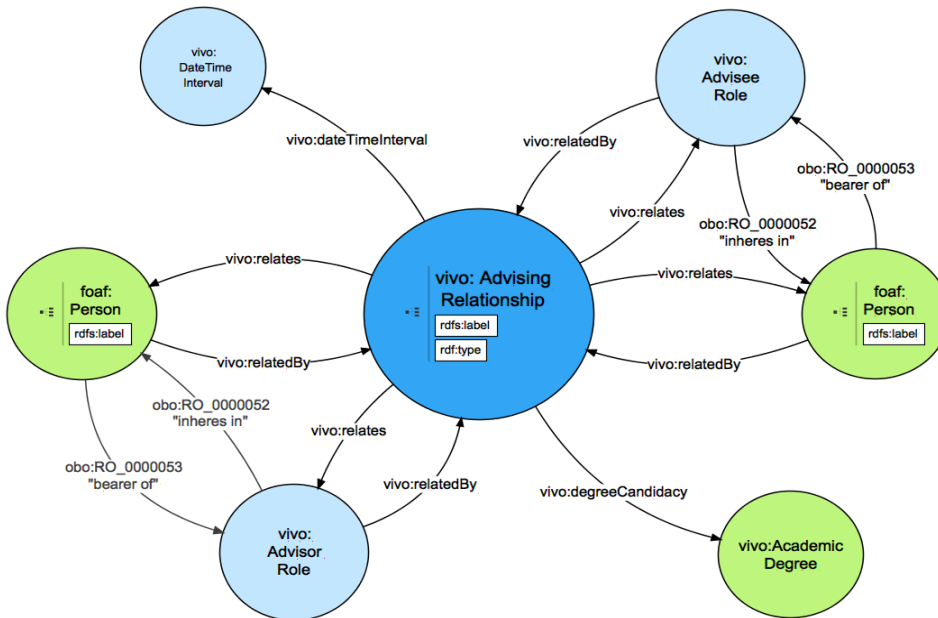


Teaching Model 12 October 2016



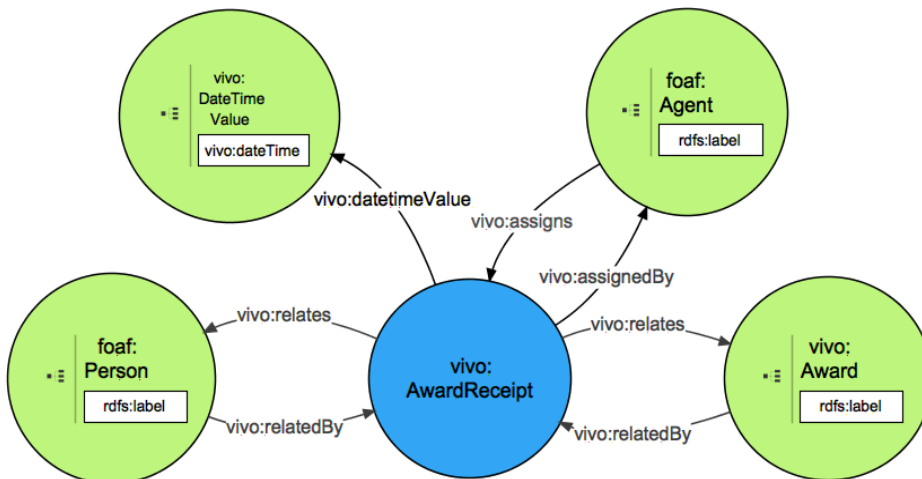
<https://wiki.lyrasis.org/display/VIVODOC111x/Teaching+Model>

Advising Model 12 October 2016



<https://wiki.lyrasis.org/display/VIVODOC111x/Advising+Model>

Award Model 15 April 2018



<https://wiki.lyrasis.org/display/VIVODOC111x/Award+Model>