

### 3

## Arquitetura de software para catalogação automática de dados geográficos

Este capítulo apresenta uma arquitetura de software para catalogação automática de dados geográficos. A seção 3.1 apresenta uma estratégia para geração automática de metadados. A seção 3.3 apresenta os requisitos de uma arquitetura de software, baseada em agentes, para catalogação automática. A seção 3.4 apresenta o detalhamento dos componentes responsáveis pela geração da descrição. Por fim, a seção 3.5 apresenta o esquema de metadados utilizado para armazenamento de metadados e os componentes responsáveis por distribuir essa descrição pelo esquema.

### 3.1. Estratégia de catalogação

Propomos uma estratégia que combina um dicionário geográfico com um catálogo de metadados. Mostraremos como utilizar o tesouro do dicionário para classificar dados geográficos e como utilizar entradas do dicionário para descrever dados geográficos.

Seja  $\mathcal{GA}$  um dicionário geográfico e assumamos que cada entrada  $\mathcal{E}$  em  $\mathcal{GA}$ , representando um objeto geográfico  $\mathcal{F}$ , tem uma representação geo-referenciada  $geo(\mathcal{E})$  de uma localização de  $\mathcal{F}$  e um  $type(\mathcal{E})$  para  $\mathcal{E}$ , cujo valor é um termo extraído do tesouro  $\mathcal{S}[\mathcal{GA}]$ .

Seja  $\mathcal{MC}$  um catálogo de metadados geográficos e assumamos que cada entrada  $\mathcal{C}$  em  $\mathcal{MC}$ , representando um dado geográfico  $\mathcal{R}$ , possui uma representação geo-referenciada  $geo(\mathcal{C})$  da área coberta por  $\mathcal{R}$  e uma descrição  $scale(\mathcal{C})$  da escala de  $geo(\mathcal{C})$ . Argumentamos que:

Q1.  $\mathcal{S}[\mathcal{GA}]$  pode ser estendido para prover também uma classificação para entradas de catálogo, isto é, para prover um  $type(\mathcal{C})$  para  $\mathcal{C}$ .

Q2. Uma descrição  $desc(C)$  de  $C$  pode ser gerada relacionando-se  $C$  às entradas do dicionário de maneira adequada.

Intuitivamente, assumindo que um dado geográfico  $\mathcal{R}$  cobre uma área na superfície terrestre, podemos descrever  $\mathcal{R}$  por uma coleção de objetos que ocorrem nessa área. Utilizamos somente os objetos cujos tipos são consistentes com a interpretação pretendida de  $\mathcal{R}$  ou cujas extensões são menores do que a escala de  $\mathcal{R}$ . Além disso, o tipo do objeto deve fornecer suficiente indicação se a escala do objeto é compatível com a escala de  $\mathcal{R}$ .

Por exemplo, seja  $\mathcal{R}$  um dado. Se  $\mathcal{R}$  deve ser interpretado como um mapa político, então  $\mathcal{R}$  deve ser relacionado com cidades e divisões políticas de uma dada área. Dependendo da escala do mapa, somente cidades acima de determinada população devem, na verdade, ser incluídas. Se  $\mathcal{R}$  deve ser interpretado como um mapa hidrográfico relacionamo-no com rios e riachos. Mais ainda, se o mapa possui uma grande escala, utilizamos somente os rios e suprimimos os riachos. Como um terceiro exemplo, se  $\mathcal{R}$  é uma imagem de satélite de uma dada área, então  $\mathcal{R}$  potencialmente representa todos os objetos que ocorrem na área. Entretanto, objetos cujas extensões são menores do que a resolução de  $\mathcal{R}$ , não devem ser relacionados a  $\mathcal{R}$ .

Em geral, sugerimos:

- a. classificar os dados geográficos também utilizando  $\mathfrak{S}[\mathcal{GA}]$ ;
- b. indicar a escala que é compatível com cada termo em  $\mathfrak{S}[\mathcal{GA}]$ . Isso é formalizado com uma função  $s: \mathfrak{S}[\mathcal{GA}] \rightarrow \mathcal{R}^*$  que mapeia cada termo de  $\mathfrak{S}[\mathcal{GA}]$  em um número Real não negativo. Para cada  $t \in \mathfrak{S}[\mathcal{GA}]$ , se  $s(t) > 0$ , interpretamos  $s(t) = n$  como indicação de que todos os objetos do tipo  $t$  são compatíveis com a escala  $1:n$ , ou menores (no sentido de que podem ser representadas naquela escala). Se  $s(t) = 0$ , interpretamos  $s(t)$  como indicação de que  $t$  pode ser utilizada para classificar dados geográficos.

Como um exemplo, considere o fragmento do *ADL Feature Type Thesaurus* mostrado na Figura 8 (a). Podemos então definir:

- $S(\text{"creek"})=10.000$  para indicar que objetos do tipo "creek" devem somente ser representados em escalas 1:10.000 ou menores;
- $S(\text{"hydrographic features"})=0$  para indicar que o termo "hydrographic features" podem ser utilizados para classificar dados.

Em certas situações, a função  $s: \mathcal{S}[\mathcal{GA}] \rightarrow \mathcal{R}^*$  pode ser parcialmente calculada a partir de outros atributos de termos do tesouro do dicionário. Por exemplo, se cada termo  $t$  sob "streams" possui um atributo  $w$  indicando a largura do rio que é classificado como  $t$ , então o valor  $w$  para  $t$  pode ser utilizado para definir  $s(t)$ .

Para a questão  $\mathcal{Q}2$ , primeiro definimos que uma entrada  $\mathcal{E}$  do dicionário, representando um objeto geográfico  $\mathcal{F}$ , é relevante para uma entrada  $\mathcal{C}$  do catálogo, representando um dado  $\mathcal{R}$ , se e somente se:

- $geo(\mathcal{E})$  e  $geo(\mathcal{C})$  são relacionadas por um dos relacionamentos topológicos usuais – toca, dentro, cruza, superpõe-se [1];
- $Type(\mathcal{E})$  é compatível com  $scale(\mathcal{C})$ .

Definimos então  $desc(\mathcal{C})$  como o conjunto de pares  $(\mathcal{E}, r)$  tais que  $\mathcal{E}$  é relevante à  $\mathcal{C}$  e  $geo(\mathcal{E})$  e  $geo(\mathcal{C})$  são relacionados por um relacionamento topológico  $r$ .

Alguns dicionários geográficos também incluem o conceito de lugar conhecido, tal como uma atração turística, uma importante cidade, etc. [13]. Se o dicionário adotado implementa esse conceito, podemos expandir a noção relevância previamente definida para incluir lugares famosos como uma informação importante. Por exemplo, considere uma imagem de satélite da cidade de Friburgo, a qual se situa a noroeste da cidade do Rio de Janeiro.

Então, em vez de somente associar a imagem com a cidade de Friburgo, podemos também indicar que a imagem cobre uma área a noroeste da cidade do Rio de Janeiro, que é um lugar conhecido. Note que ao relacionarmos lugares conhecidos com recursos de informação, adotamos relacionamentos direcionais – a norte de, a sul de, a leste de, etc. – ou relacionamentos qualitativos, tal como perto.

Como um exemplo da geração de  $\mathit{desc}(C)$ , suponha que adotamos o ADL Gazetteer e o ADL Feature Type Thesaurus. Considere um fragmento de imagem da cidade do Rio de Janeiro, retirada do website Brasil visto do espaço<sup>13</sup> [14], mostrado na Figura 8.

Essa imagem será processada da seguinte forma:

1. Extrair os parâmetros de geo-referenciamento do dado geográfico. Nesse caso, o fragmento de imagem é consistente com a escala 1:25.000 e tem como retângulo de borda definido pelo par de coordenadas ((43° 15' W, 22° 52' 30" S), (43° 07' 30" W, 23° S)).
2. Assuma que o usuário escolha relacionar o fragmento de imagem com “*hydrographic features*”, um termo do ADL FTT que, em nosso exemplo, pode ser usado para classificar dados geográficos.
3. Como as entradas do ADL Gazetteer não têm nenhuma informação de escala, ela será ignorada.
4. Acesse o ADL Gazetteer, usando os parâmetros extraídos no passo 1 e os termos ADL FTT sob “*hydrographic features*”, selecionado no passo 2. A consulta retornará nove entradas, das quais as três primeiras são:
  - 4.1. Objeto (“Rodrigo de Freitas, Lagoa – Brazil”, “*lakes*”, “*within*”)
  - 4.2. Objeto (“Comprido, Rio – Brazil”, “*streams*”, “*within*”)
  - 4.3. Objeto (“Maracanã, Rio – Brazil”, “*streams*”, “*within*”)
5. Armazene o resultado da consulta como uma descrição do dado, isto é, como uma lista de pares  $(\mathcal{N}, r)$ , onde  $\mathcal{N}$  é o objeto geográfico recuperado no passo 4 e  $r$  é o relacionamento topológico entre a imagem e  $\mathcal{N}$  (nesse exemplo  $r$  é “*within*”).

---

<sup>13</sup> <http://www.cdbrasil.cnpm.embrapa.br/>

Esse breve exemplo ilustra algumas das idéias básicas desse trabalho. Primeiro, o uso do dicionário para também classificar dados geográficos, evita adotar um segundo esquema de classificação, tal como, ISO19115 Topic Categories [10]. Entretanto, requer a definição de uma função de compatibilidade  $s: \mathcal{S}[GA] \rightarrow \mathcal{R}^*$ .

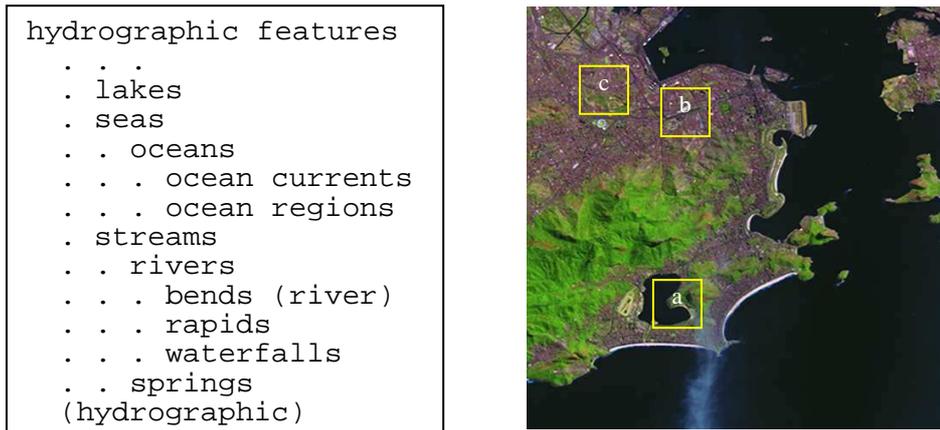


Figura 8 – Um fragmento do ADL Feature Type Thesaurus e uma imagem da cidade do Rio de Janeiro

Segundo, uma descrição útil de um dado geográfico  $\mathcal{R}$  pode ser criada como uma lista de pares  $(\mathcal{N}, r)$ , onde  $\mathcal{N}$  é um objeto geográfico e  $r$  é o relacionamento topológico entre  $\mathcal{R}$  e  $\mathcal{N}$ , obtido pela consulta ao dicionário. Além disso, a lista contém somente objetos cujos tipos são compatíveis com a escala de  $\mathcal{R}$ .

### 3.2. Requisitos da arquitetura

Para compreendermos o processo de catalogação automática e identificarmos os requisitos de software necessários, consideremos a situação de uma empresa que recebe, periodicamente, imagens de satélite e deseja armazená-las de modo a poder selecionar aquelas que são adequadas a um projeto.

Freqüentemente, imagens de satélite são gravadas em arquivos utilizando-se padrões de compactação como, por exemplo, o GeoTIFF que permitem

anexar aos dados da imagem, contidos no arquivo, as coordenadas geográficas da região na superfície terrestre coberta pela imagem.

Tais imagens podem ser obtidas de várias fontes e serem utilizadas em vários setores da empresa. Por exemplo, poder-se-ia obter imagens da EMBRAPA<sup>14</sup> e do LANDSAT<sup>15</sup> para um departamento que desejasse analisar recursos hidrográficos. Outro departamento, no entanto, poderia estar interessado nos municípios cobertos pelas mesmas imagens.

Podemos, então, conforme a seção 3.1, definir um algoritmo para catalogação automática de dados geográficos:

catalogação de dados geográficos  
*enquanto* (interrupção não for solicitada) *faça*

- Obter novo dado;
- se houver novo dado *então*
  - Extrair metadados do seu conteúdo;
  - Criar descrição parcial;
  - Classificar dado;
  - para cada* dicionário geográfico configurado *faça*
    - Buscar metadados relacionados com o dado;
    - Eliminar metadados incompatíveis com escala e classificação do dado;
    - Acrescentar metadados à descrição;
  - se houverem metadados provenientes dos dicionários *então*
    - Catalogar dado;

Para implementar o algoritmo propomos dois agentes de software. O primeiro deles, o agente rastreador (*CrawlingAgent*), é responsável por inspecionar repositórios de dados, buscando novos dados a serem catalogados. Uma vez encontrados, o agente extrai metadados do seu recipiente<sup>16</sup> e, em seguida, envia uma mensagem a outro agente, o “agente catalogador”,

---

<sup>14</sup> <http://www.cdbrasil.cnpm.embrapa.br>

<sup>15</sup> <http://landsat.gsfc.nasa.gov>

<sup>16</sup> Recipiente de um dado é a sua unidade elementar de armazenamento. Dados são freqüentemente armazenados em arquivos ou registros de bancos de dados.

solicitando a catalogação da imagem. Não limitamos o “agente rastreador” a somente uma instância. Podem ocorrer situações em que várias instâncias independentes estejam ativas para garantir robustez e flexibilidade ao processo. Poderíamos, por exemplo, configurar um agente para buscar as imagens geradas pela EMBRAPA e outro para as do LANDSAT.

O agente catalogador (*CataloguingAgent*) é o agente responsável por buscar, em um dicionário geográfico, objetos relacionados ao dado geográfico e, caso seja bem sucedido, armazená-los no catálogo. Periodicamente, o agente verifica, em um canal de comunicação comum, se existem novos dados a serem catalogados. Em caso afirmativo, recupera suas coordenadas geográficas e extrai de dicionários objetos relacionados às coordenadas. Com base nos dados dos dicionários, cria descrições dos dados e os armazena no catálogo. Esperamos que um processo de catalogação manipule somente um catálogo, portanto, sugerimos somente uma instância do “agente catalogador”.

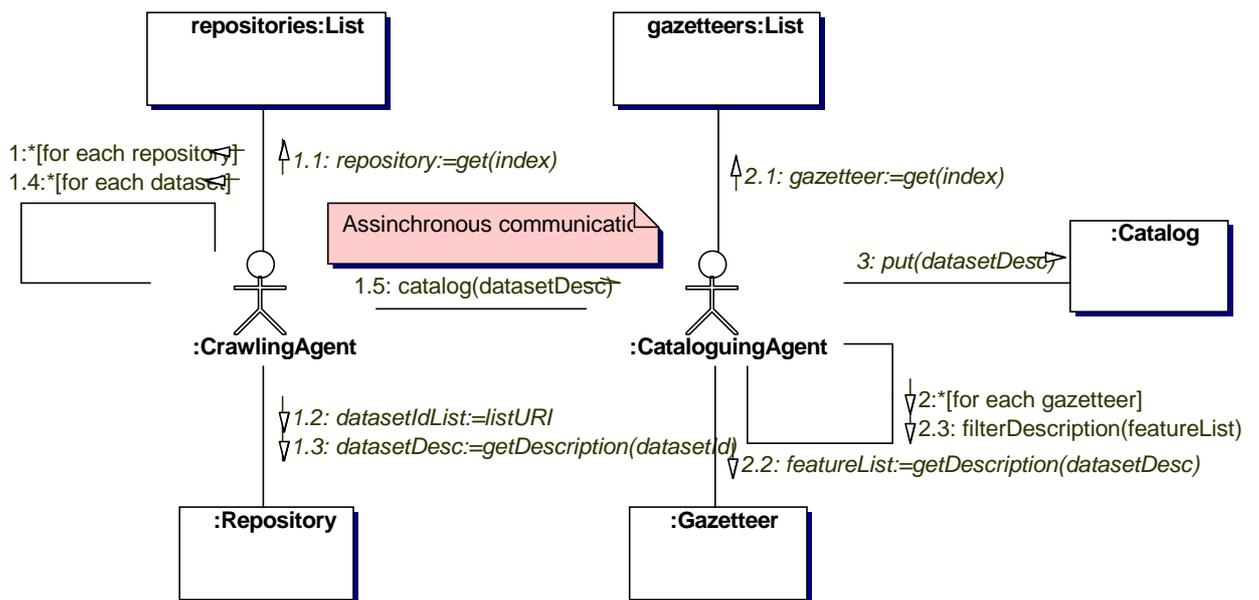


Figura 9 - Arquitetura de agentes para o processo de catalogação

Outros catálogos podem ser alimentados utilizando-se outros processos de catalogação, cada qual com seu próprio conjunto de agentes e configurações específicas. Um esquema dessa arquitetura é mostrado na Figura 9.

Os repositórios (*Repository*), na figura, são os locais nos quais se encontram novos recipientes de dados geográficos. Note que não está

representado no esquema o processo para captura desses dados do provedor. Por exemplo, se a EMBRAPA disponibiliza as imagens em algum endereço na internet e os arquivos são transferidos para o cliente, esse procedimento não está representado. O local em questão é o local final dos arquivos de imagens que os agentes irão consultar.

Os dicionários (*Dictionary*) são os dicionários geográficos dos quais serão obtidos objetos geográficos relacionados ao dado.

Pelo exposto até aqui podemos listar os seguintes requisitos funcionais para um software de catalogação automática:

1. Permitir a criação de processos de catalogação informando: repositórios, dicionários, catálogo e um identificador para o processo de catalogação;
2. Permitir a alteração das configurações dos processos de catalogação;
3. Permitir remover configurações de processos de catalogação;
4. Permitir armazenar as configurações dos processos de catalogação;
5. Permitir iniciar e interromper cada processo independentemente;
6. Exibir uma lista de falhas ocorridas nos processos de catalogação;
7. Exibir o estado dos processos de catalogação;

Propomos, ainda, alguns requisitos não funcionais para flexibilizar o processo como um todo e garantir o bom desempenho dos equipamentos hospedeiros dos agentes.

Uma atenção especial deve ser dedicada à questão do desempenho. A natureza da operação de catalogação pode produzir grande carga de processamento dos agentes. Para minimizar os efeitos do consumo excessivo de recursos do ambiente computacional os agentes devem ser capazes de interromper seu processamento por alguns instantes em intervalos regulares de tempo ou mesmo identificar períodos de ociosidade dos recursos e aproveitá-los.

Um formato de imagem frequentemente utilizado na geração de imagens geográficas é o GeoTIFF, como já mencionado. Entretanto, podem ocorrer situações onde outros formatos para codificação de dados geográficos são

utilizados, por exemplo o ESRI Shapefile. É desejável que o software possa ser facilmente adaptado a novos padrões de arquivos contendo dados geográficos.

Na arquitetura apresentada anteriormente os Agentes Rastreadores têm o papel de procurar por novos dados em repositórios previamente especificados no processo de catalogação. Tais repositórios podem estar disponíveis sob vários protocolos. Poderiam ser diretórios no sistema de arquivos, *sites* FTP, URL's acessadas através de conexões HTTP, ou outro tipo de repositório qualquer. Deseja-se flexibilizar os tipos de repositórios com os quais o agente pode se comunicar.

Finalmente, analogamente aos repositórios de dados, o software deveria poder se adaptar a diferentes dicionários, com protocolos de comunicação diferentes.

Note que esses três últimos requisitos não se tratam de adaptação dinâmica, mas sim de conferir ao software uma arquitetura flexível o suficiente de modo que possa ser facilmente modificado para se compatibilizar com os novos componentes externos.

### **3.3. Processo de catalogação**

A Figura 10 apresenta um esquema dos componentes responsáveis pelo processo de catalogação.

Podemos definir, de acordo com a seção 3.2, que um processo de catalogação (*CataloguingProcess*) possui as seguintes propriedades: nome (*name*), estado (*state*), coleção de repositórios (*repositories*), coleção de dicionários (*dictionaries*), catálogo (*catalog*) e repositório de falhas (*fails*).

Ao iniciar-se a execução de um processo de catalogação (operação *start()*), os agentes são criados e configurados. Os repositórios podem ser distribuídos entre vários agentes catalogadores, no entanto, cada agente catalogador deve consultar todos os dicionários. A cooperação entre eles ocorre na concorrência para catalogar os novos dados geográficos identificados. Para podermos distribuir a consulta aos dicionários entre vários agentes seria preciso definir

regras de interação entre os catalogadores de modo a garantir que, para cada novo dado todos os dicionários seriam consultados e que seria possível identificar quando isso teria ocorrido, permitindo, assim, a catalogação do dado. Essa coordenação, no entanto, não será tratada nesse trabalho.

Ao menos um agente de cada tipo deve ser criado, mas a arquitetura não limita o número máximo. Há que se estudar cada situação para optar por um conjunto adequado de agentes e implementar o algoritmo apropriado. No protótipo foi implementado apenas um agente de cada tipo.

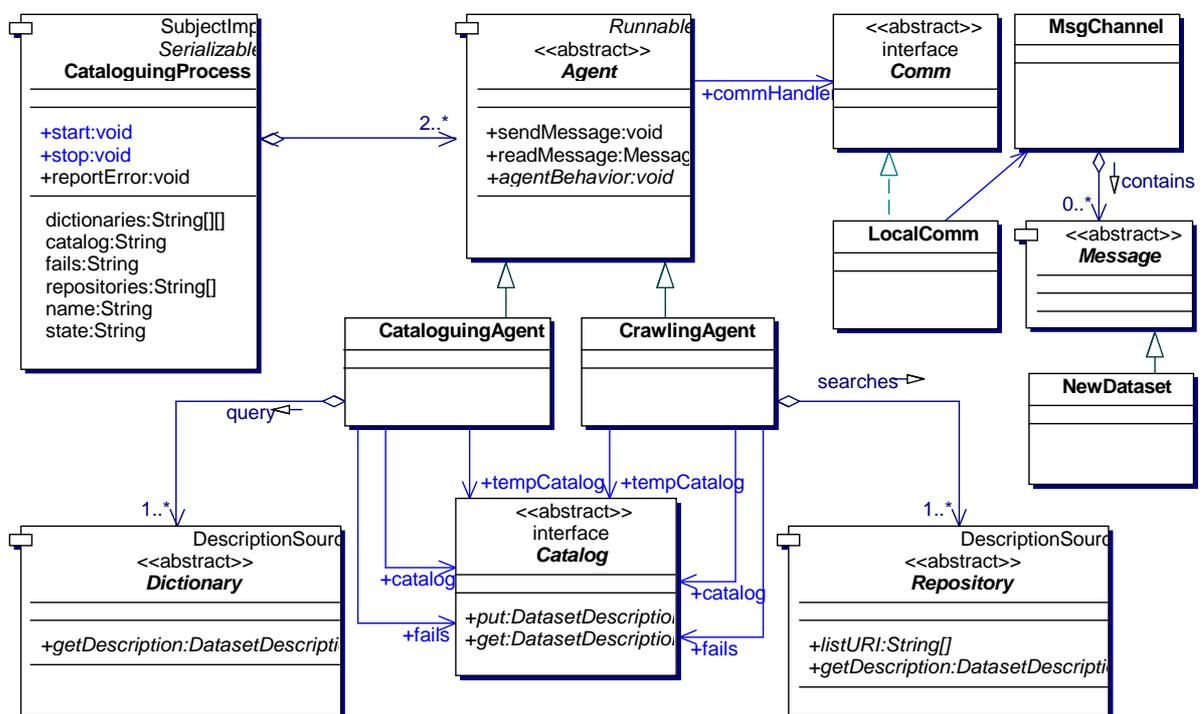


Figura 10 - Processo de catalogação automática

Um catálogo temporário comum a todos os agentes serve para identificar os dados que ainda não foram catalogados, mas já foram carregados, evitando-se o seu reprocessamento.

Cada agente rastreador solicita em intervalos regulares de tempo a lista dos identificadores de recipientes de dados existentes nos repositórios (operação *Repository->listURI()*). Cada recipiente contém um dado geográfico a ser descrito e catalogado. O identificador do recipiente assumirá o papel de identificador da descrição do dado no sistema. Os identificadores que não

existirem no catálogo, nem no repositório temporário ou de falhas são considerados novos e os metadados do conteúdo dos respectivos recipientes são capturados (operação *Repository->getDescription(String uri)*). Uma descrição parcial do dado é criada, a partir dos metadados capturados, e enviada (operação *CrawlingAgent->sendMessage(Message message)*) ao agente catalogador em uma mensagem do tipo “novo dado” (*NewDataset*). Entre outras informações extraídas do conteúdo do arquivo estão as coordenadas geográficas da região na superfície terrestre coberta pela imagem.

Cada agente catalogador solicita continuamente ao canal de comunicação uma mensagem do tipo “novo dado” (operação *CataloguingAgent->readMessage(new NewDataset( ))*). Havendo mensagem, o agente consulta cada um dos dicionários para recuperar objetos geográficos contidos na área coberta pela imagem. Não ocorrendo nenhuma falha, os objetos são incluídos na descrição do dado e esta é incluída no catálogo. Caso não exista nenhuma mensagem o agente interrompe o processamento e aguarda até que uma nova mensagem seja enviada.

Uma falha ocorre quando não é possível recuperar nenhuma informação sobre um dado geográfico de um dicionário. Nesse caso, o dado é armazenado no repositório de falhas.

Erros fatais detectados durante o processamento de qualquer agente são imediatamente reportados ao processo de catalogação (operação *CataloguingProcess->reportError( )*) que interrompe a execução de todos os agentes do processo (operação *CataloguingProcess->stopAgents( )*) e notifica os observadores do processo (operação *CataloguingProcess->notifyObservers( )*) do ancestral *SubjectImpl* de *CataloguingProcess* – vide padrão Observer na seção 4.3). Alguns exemplos de erros são: endereço de serviço do dicionário inválido, repositório inválido, falha na comunicação de rede, catálogos inválidos, etc.

### **3.4. Geração de metadados**

A Figura 11 apresenta o esquema dos componentes envolvidos na geração automática de metadados.

Existem dois tipos de fontes de descrição (*DescriptionSource*) de dados geográficos: o próprio recipiente do dado em um repositório (*Repository*) e dicionários geográficos (*Dictionary*). O procedimento de captura da descrição baseia-se em dois elementos abstratos: decodificador (*Parser*) e construtor (*Builder*).

O decodificador é responsável por identificar os elementos de metadados tanto no recipiente do dado, quanto na resposta de uma consulta a um dicionário. Assumiremos, para facilitar o entendimento no momento, que tanto os metadados armazenados nos recipientes dos dados geográficos, quanto aqueles obtidos de um dicionário geográfico serão codificados em XML. O componente *XMLParser*, extensão do decodificador, implementa um algoritmo para identificação dos elementos XML. Caso fosse necessário interpretar metadados armazenados no formato GeoTIFF poderíamos criar um *GeoTIFFParser* que implementasse o algoritmo apropriado.

Uma fonte de descrição pode possuir vários decodificadores de metadados. A escolha apropriada do decodificador, na geração de uma descrição, depende da identificação do formato no qual os metadados serão obtidos. Com a possibilidade de termos várias estratégias diferentes de decodificação, pela extensão do componente decodificador, obtemos flexibilidade na variedade recipientes e dicionários. Esse ponto é fundamental para que, como veremos na seção 4.6, possamos generalizar o mecanismo de catalogação automática para dados não geográficos.

Um decodificador transforma os metadados obtidos de uma fonte de descrição em uma coleção de elementos de metadados (*MetadataElement*). Para informações codificadas em formato XML cada elemento de metadado representa uma *tag* com seu nome, valor e atributos e valores de atributos. A coleção também poderia representar metadados em recipiente GeoTIFF. Nesse caso, os atributos assumiriam valor nulo.

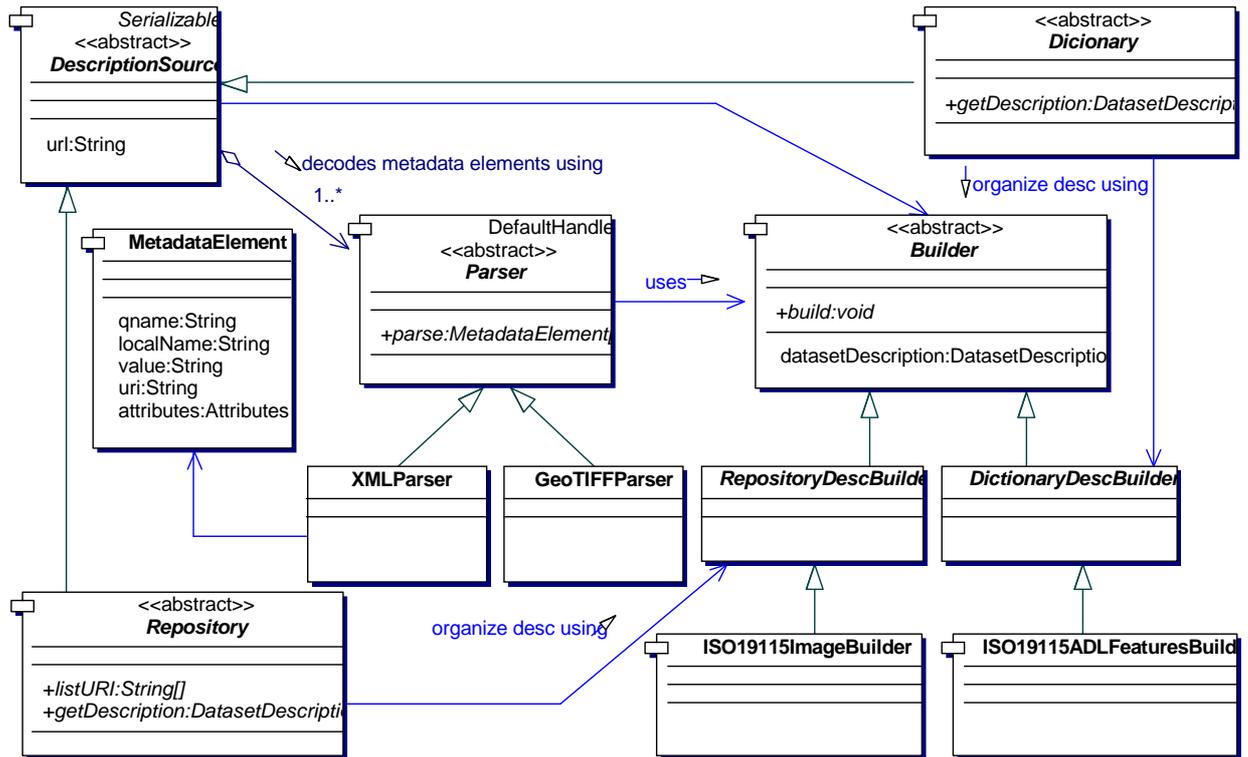


Figura 11 - Recuperação de informações

O segundo componente no processo de geração de metadados, o construtor, é responsável por receber uma coleção de elementos de metadados e os organizar em um esquema de metadados específico. Caso os metadados estejam sendo recuperados de um recipiente de dado geográfico, um construtor de descrição de recipiente (*RepositoryDescBuilder*) processa a coleção. Se estamos recuperando metadados de um dicionário geográfico, um construtor de descrição de dicionário (*DictionaryDescBuilder*) se encarrega da organização.

Cada decodificador de fonte de descrição possui uma referência ao construtor associado à essa fonte, isto é, um *XMLParser* associado a um repositório possui um *RepositoryDescBuilder* e um *XMLParser* associado a um dicionário contém um *DictionaryDescBuilder*.

Pode ocorrer que se deseje organizar as informações em mais de um tipo de esquema. Nesse caso, haveriam diferentes especializações do construtor. A Figura 11 apresenta construtores que utilizam somente o esquema de metadados definido pela ISO 19115, conforme exposto na seção 2.1.1. O componente *ISO19115ImageBuilder* organiza metadados de recipientes e o

*ISO19115ADLFeaturesBuilder* organiza metadados a partir de objetos recuperadas do ADL Gazetteer.

Para obter metadados em um recipiente o agente rastreador informa ao repositório o identificador do recipiente (operação *Repository->getDescription(String identifier)*). O repositório identifica o formato de codificação do recipiente, escolhe o decodificador apropriado, e solicita a ele a decodificação do conteúdo do recipiente (operação *XMLParser->parse(InputSource source)*). O decodificador, ao final do seu processamento, solicita ao construtor para organizar os elementos de metadados (operação *RepositoryDescBuilder->build(MetadataElement[] metadataElements)*). Ao final, o repositório solicita ao seu construtor a descrição resultante (operação *RepositoryDescBuilder->getDatasetDescription()*).

Para obter metadados em um dicionário o agente catalogador informa ao dicionário a descrição parcial do dado obtida do seu recipiente (operação *ADL Gazetteer-> getDescription(DatasetDescription datasetDescription)*). O dicionário obtém da descrição as informações relativas às coordenadas geográficas e faz uma consulta ao repositório físico do ADL Gazetteer que recupere todos os objetos contidos na área geográfica do dado. A resposta da consulta é decodificada (operação *XMLParser-> parse(InputSource source)*) e enviada ao construtor para organização (operação *DictionaryDescBuilder-> build(MetadataElement[] metadataElements)*). Ao final, o dicionário solicita ao seu construtor a descrição resultante (operação *DictionaryDescBuilder-> getDatasetDescription()*).

### 3.5. Esquema de metadados

A Figura 12 apresenta o esquema de metadados utilizado para armazenar a descrição de um dado geográfico e os componentes responsáveis por distribuir essa descrição pelo esquema.

Os metadados de um dado geográfico são organizados segundo um esquema conceitual determinado pela norma ISO19115. Segundo ela, um dado geográfico (*DS\_Dataset*) pode ser uma imagem, um mapa ou até mesmo um simples objeto geográfico. Cada dado está associado a um conjunto de

elementos de metadado (*MD\_Metadata*) que o descreve. A norma unifica, portanto, os conceitos de dado geográfico e objeto geográfico utilizados na arquitetura de catalogação. Introduce, ainda, o conceito de agregação de dados (*DS\_Aggregate*) que correlaciona dados geográficos afins. No nosso caso, definiremos que correlaciona o dado geográfico com os objetos geográficos obtidos dos dicionários. O agregado pode, também, estar associado a um conjunto de elementos de metadados para sua descrição.

Foi utilizado, apenas, um subconjunto das estruturas de metadados definidas pela norma, suficiente para a compreensão da arquitetura, são elas: *DS\_Dataset*, *DS\_Aggregate*, *DS\_OtherAggregate*, *MD\_Metadata*, *MD\_SpatialRepresentation*, *MD\_gridSpatialRepresentatin*, *MD\_Georectified* e *GM\_Point*. Os componentes que não pertencem à ISO 19115 e apresentados no esquema são responsáveis por capturar os metadados e distribuí-los pelas estruturas da norma. Essa estratégia permite flexibilidade na alteração do esquema de metadados e até mesmo da forma de armazenamento. Podemos armazenar os metadados descritos pela ISO em um banco de dados orientado à objeto ou como instâncias de uma ontologia.

As estruturas de dados definidas pela ISO 19115 são manipuladas por meio do componente *OOISO19115GeoImage*. Suas operações fornecem ao software de catalogação uma interface para armazenamento dos metadados. O componente *OntISO19115GeoImage* foi apresentado somente como ilustração de variações possíveis para armazenamento de metadados. Com ele pretendemos sugerir que os conceitos ISO poderiam ser traduzidos para uma representação em forma de ontologia permitindo armazenar os metadados em bancos de dados semânticos. A generalização desses dois componentes até *DatasetDescription* pretende apoiar o *framework*, de que falamos, para catalogação de dados não geográficos.

