

# **Automated Data Exchange to and from the National Data Repository in the Netherlands\***

**Stephan Gruijters<sup>1</sup> and Sjaak Derksen<sup>1</sup>**

Search and Discovery Article #70337 (2018)\*\*

Posted May 14, 2018

\*Adapted from extended abstract based on oral presentation given at the GEO 2018 13th Middle East Geosciences Conference and Exhibition, March 5-8, 2018, Manama, Bahrain

\*\*Datapages © 2018. Serial rights given by author. For all other rights contact author directly.

<sup>1</sup>TNO Geological survey of the Netherlands, Utrecht, NL ([stephan.gruijters@tno.nl](mailto:stephan.gruijters@tno.nl))

## **Abstract**

The exploration and production for oil and gas resources produces a vast volume of data (seismic surveys, borehole logs, well reports, production data, license data etc.). This information is stored in a National Data Repository (NDR). Reporting requirements, whether or not governed by law, steer the data exchange between operators and the NDR. Access to the information in the NDR's varies significantly around the world. In some countries, the information is only available to a distinct group of (member) organisations in data rooms, where in other countries the information is publicly available via a web interface, sometimes, even without charge. The Dutch NDR not only holds data related to exploration and production of energy resources, but also data related to groundwater and the shallow subsurface collected during the last century. Furthermore, Geo(hydro)logical models created from these data are stored in the same NDR. The information is free for anyone to use (open data policy) via internet (DINOloket.nl). The importance of this database has been recognized by the Dutch government, leading to the upgrade of this system to a formal 'key' register governed by Law: 'Basis Registratie Ondergrond' (BRO). The main driver behind the BRO is improving the reliability of subsurface information, diminishing costs (single entry, multiple use) and reduce risks in (large scale) (infra)structure projects. This BRO in principle is a system that connects to systems of data providers (via a SOAP intake service 'is') and data users (via a SOAP dispatch service 'ds'). Since 2010, TNO is developing and building this key register. The register covers 26 data types within 6 data domains. The data models and exchange formats used are being developed with input from data providers and users. Business rules in the web services guarantee consistency to the data model, while the owner of the data has to make sure the data is correct. Since January 1, 2018 the first three data types are live (technical system ready and legislation effective). This not only enables data providers and users to create a live connection between the key register and their own system, it also opens a completely new arena for software developers to create (commercial) added value applications on top of the database. We believe that this will increase re-use of data tremendously and stimulate new economic activities related to (big) data exploitation: the new trend for the coming decade.

## **eGovernment**

Since the late nineties the use of ICT to exchange information between governmental organisations, civilians and companies has intensified. In The Netherlands this has accumulated to a set of 11 formal 'key' registers ('basis registraties') [[www.digitaleoverheid.nl](http://www.digitaleoverheid.nl)]. The rationale behind

this system is that the public government has to be informed ('We know what we need to know'), knows correct from incorrect ('We can trust our information system'), is cost effective and service oriented ('We ask for information only once, re-use it multiple times'). For the key register on the subsurface, an additional business case exists: reduction of failure costs in (large-scale) infrastructure and construction projects. Very often during the planning and design phase of such projects, the access to reliable, comprehensive sub surface data is sub optimal. This may result in poor designs leading to failure (flooding, cracks, differential settlement, and even collapse) of the (infra)structure during the construction phase or even later, when the structure is already in use. The key register of the subsurface will improve the understanding of the geological structure and geotechnical properties of the subsurface, giving geologists and engineers a better starting point. In addition, because of the mandatory delivery of newly acquired data, the amount of data inside the key register will grow substantially every year, further increasing the quality of this starting point. The set of key registers hold information on people (identification and income), companies, cars, real estate, buildings (dimensions and economic value), addresses, topography (detailed - up to 1:500 scale and less detailed – from 1:10.000 to 1:1.000.000 scale) and finally the subsurface. An important concept for the information in the key registers is the distinction between 'non-authentic' and 'authentic' data. The latter can be used directly by the government without further investigation. To improve the reliability of the authentic data all key registers are designed to be self-repairing: they have a (mandatory) reporting mechanism'. If a data consumer has reasonable doubt on the correctness of the data, he can (governmental organisations have to) report this to the owner of the data. The owner needs to conduct an investigation to check if the authentic data needs to be corrected. In the mean time, the data object in the register is marked as 'under investigation' to inform data users that this specific entry does not have to be used.

### **BRO Development**

Key registers are typically embedded in legislation. In case of the BRO the first thought of having a key register on the subsurface dates from 2002 when the Geological Survey of the Netherlands discussed the possibility to 'upgrade' the existing National Data Repository (Data en Informatie van de Nederlandse Ondergrond, DINO) with the ministry of spatial planning and the environment. It took more than 15 years (see [Figure 1](#)) to have a law on the Key Register of the subsurface (accepted by parliament on 30-09-2015). This law is implemented with additional regulations that specify the data standard Catalogue ('Algemene maatregel van Bestuur', AmvB) and the details for mandatory use ('Koninklijk Besluit', KB). Starting in 2009 the technical system that needs to facilitate this Key Register was designed, built and tested. In the beginning (2009) a classic waterfall method was used. This started with a system architecture, requirements and use cases. Using these as a blue print, the (detailed) specifications were written, code was developed and the software was tested before it was released. In 2015, this classic approach was more and more mixed with agile/scrum methodologies. First only within the code development, expanding later to incorporate testing. Finally, the detailed functional specifications were replaced by user stories (as ... I would like to be able to ... to make sure that...). The released software was tested with stakeholders in pilots to see how the BRO fits the existing business processes in practice. For the first three RO's (see next paragraph for further explanation) a number of pilots were conducted: five for CPT and two for GMW. The BHR web service was developed in close cooperation with Wageningen Environmental Research (Alterra), the organisation that produces 95% of this data, effectively making a pilot obsolete. Close to the actual go-live date an end-to-end test was performed, testing the entire information chain from production of new data in the field up to downloading this data from the BRO.

## BRO Data and standardisation

The BRO covers 26 data types (referred to as registration objects or RO's) from six data domains: (1) soil and subsurface investigations, (2) groundwater monitoring, (3) soil quality monitoring, (4) the Mining law, (5) groundwater use and (6) subsurface models (see [Figure 2](#)). For each RO a data model is constructed. This is done in close cooperation with stakeholders and experts from the domain who understand the data itself as well as the business processes being used during the acquisition and processing of the data. During this standardisation process, the data model is published on internet periodically for public review, making it fully transparent and giving every stakeholder the opportunity to provide feedback. In the end, a formal document (catalogue) is produced describing the context of the RO and the semantic data model (entities, attributes, business rules) in detail.

The 6 data domains differ in characteristics, influencing the standardisation process. Some of them have a vast group of governmental organisations involved, e.g. cone penetration tests are used by almost all governmental organisations in the Netherlands (at least 423: 377 municipalities, 12 provinces, 22 waterboards, 12 ministries). This results in a large number of stakeholder organisations, each having their own list of requirements for the standard that needs to be developed. That is why the standardisation process focuses on the data that are relevant for re-use by third parties instead and leaves the data that are only relevant for a limited group of stakeholders to be managed in their own systems. The mining law domain on the other hand has mainly the ministry of economic affairs as dominant organisation. The E&P sector is international and already has recognised the importance of standardisation, resulting in a set of already existing (commercial) standards for data storage and exchange. The challenge here is to see whether these standards can be (partly) re-used to minimize the adjustments the industry has to make without jeopardizing the re-usability of the data by other domains. Furthermore, the introduction of the BRO actually makes the data types in this domain subject to two laws: the BRO law and the mining law. The main difference is that the mining law regulates the exploration and exploitation of (energy)resources in the Netherlands, and the BRO law regulates the data(exchange) formats. The two monitoring domains consist of RO's that are very much related to each other (e.g. monitoring nets consisting of monitoring points, where a monitoring well is present from which - changes in - groundwater levels and groundwater quality is measured). Each RO's has his own stakeholder community, but the benefit for re-use lies in the combination of all RO's in the domain. This challenges the standardisation team: decide what part of the information that is needed for the entire domain has to be stored in which RO. Finally the domains of models: The geological models in the BRO are constructed only by the Geological Survey of the Netherlands. Therefore, a dedicated connection between the working environment of the geologists and the BRO is to be preferred over SOAP services. This means that standardisation will focus on an exchange format for the use of the geological models by third parties.

As of January 1, 2018 three RO's are implemented (1) Cone Penetration Tests - CPT, (2) Groundwater Monitoring Well - GMW and (3) BoreHole Research – BHR. A CPT is an investigation method that measures the (point and sleeve) resistance of a sensor (cone) that is pushed into the subsurface mechanically; up to about 25 meters below surface level. It is a measurement typically used to calculate the bearing capacity and compressibility of (deltaic, fluvial or Eolic ) sediments. A GMW is the infrastructure (a (set) of screens) that is used to measure the groundwater level and or take groundwater samples at a specific depth range in the subsurface. Finally, a BHR is an investigation where a boring is used to explore the subsurface. This investigation typically consist of a borehole log describing the sediments encountered during the execution of the boring possibly expanded with other sensor data measured in the borehole and/or (chemical/physical/geotechnical) analysis from samples taken during the boring process.

The first version of this BRH consists of the borehole log for shallow borings (typically up to 1.5 meter below surface level) described using the nomenclature used in the soil domain. Based on these three RO's a template has been derived for modelling the RO's (see left panel [Figure 3](#)). This template consists of five clusters. The first cluster (Registration Object) identifies the RO and the organisations that are responsible (meaning they deliver the data to the BRO) and accountable (meaning they are accountable for the data that has been delivered) for the delivery. In practice this can be the same organisation. The second cluster (Meta data) holds a set of attributes describing the context of the RO. This cluster is different for each RO even though some attributes are similar for all three RO's. The third cluster is the location, describing the coordinates and vertical position of the RO. The first and third clusters are the same for all three RO's. The fourth cluster holds the measured data and is different for each RO. Finally, the Registration History is derived by the BRO system itself, storing all formal changes made in the data. Each RO is delivered to the BRO using a (SOAP) registration request holding the source document (capturing the data from the field) and additional data to identify the transaction itself. For the current implementation of CPT and BHR there is only one source document; all data is created in one go and delivered to the BRO. For GMW there is a set of 11 source documents. This is because a GMW is likely to change over time: new data is added to the initial situation (e.g. when a new sensor is placed in the well, or the owner of the well changes). To correct possible false information in the BRO, these same source documents are used, only now in a correction request.

The semantic data model is transferred to a technical data model (XSD and WSDL) that describes the definition of the data model and the operations of the web service (see [Figure 4](#)). The BRO uses a standard library that is common for all RO's (BRO common). Within each RO, common attributes that are used by the intake service and the dispatch service are stored in the RO-common libraries. Finally, parts that are unique for the intake service and the dispatch service are stored separately. All these libraries connect to the open GIS profiles (gml, observation and measurement, swe, sa). The semantic data models are available at <https://bro.pleio.nl/standaarden> (in Dutch); together with documentation describing the functionality of the SAOP services and technical documentation describing the technical data model (XSD/WSDL). The technical data models themselves are published at <https://schema.broservices.nl/>.

### **BRO System: Interconnectivity**

The BRO is connected to the (closed) network that interconnects Dutch government organizations (see [Figure 5](#)). This network is called 'Digikoppeling' v2.x, which is probably best translated as DigiConnection. The technology used by the BRO is called WUS, which stands for WSDL, UDDI and SOAP. These are all open standards, standardized by OASIS and W3C. Authenticity of clients is established by means of a client side certificate. The Digikoppeling network also contains a subscribe / publish mechanism for changes. Criteria (amongst other ROxyz criteria) for being notified of changes are typically a geographical area of interest, for example, an administrative zone (e.g. municipality) or a project-site (e.g. a trace for a new motorway). The BRO pushes (notifies) changes in ROxyz via ebMS2 (ebXML Messaging Service) to the Digilevering system. Digilevering then notifies interested parties that subscribe to the criteria that match the change. An interested party is then able to fetch the full ROxyz from the BRO. All the data inside the BRO is open (accessible for everyone, free of charge) There are several user groups that are served by the BRO ranging from Engineering companies, consultants, science, journalism to governmental agencies. They all have different requirements in how they want to access the BRO. Therefore, the BRO has several outlet channels. First, there is DINOLoket, a rich client website used by (but not limited to) expert users from the different domains. The user can search, inspect and download data via a map interface. The inspect functionality is implemented with views of sample photos and graphical representations of e.g. borehole logs, water tension graphs and grain size distributions. The geological models are presented as top, bottom and thickness grids and 3D voxels. They can be

inspected with user defined horizontal and vertical cross sections. This same portal used to disseminate data from the DINO system. During the transition period where data will be converted from DINO to BRO both systems will feed data to this portal. The BRO data is also published in a lighter web interface called PDOK, offering the data as WMS, WFS and Atom services. Finally, there is European ruling that also demands publishing datasets in a harmonized standards way called INSPIRE. With an ETL procedure the BRO data is mapped and transformed to the official EU mandated INSPIRE exchange standards.

### **BRO System: Functional Decomposition**

The BRO System consists of a set of eight (reusable) functional components (see [Figure 6](#)). The (1) Intake services ROxyz (ROxyz is a place holder for a Registration Object with its versioning xyz ) are responsible for registering new RO's, managing the life cycle of RO's and correcting existing data in the BRO related to these RO's. Typically, a RO has to obey a certain set of business rules associated with the Registration Object. These business rules cover internal consistency, consistency related to previously registered ROxyz data, consistency in the business process leading to the request, etc. Also, the existence of parties mentioned or present in the data is verified. As an extra service for the data owners, an intake portal is added to the system ('bronhouderportaal'). Via this portal, data owners (delivery accountable party) can manage their supply chain and select a data provider from it (delivery responsible party). The latter can upload xml files to the portal that subsequently are checked by the data owner. Only when the files are labeled 'ok', the intake portal sends them to the BRO using the intake service. The (2) Validation services work exactly on the same operations as the intake service with the exception that the data is not committed to the database. It is used by data providers to check their data before delivering it to the BRO. The (3) Dispatch services are used to extract data from the BRO. The ds has two operations. The first one selects a subset of RO's from the BRO that meet user defined selection criteria, either general (geographic criteria, or registration object history) or RO specific (depth or other parameters). It returns a summary of ROxyz characteristics including an identifier. This identifier can be used to extract all data related to this RO in the BRO. In the near future we expect an (yet unexplored) requirement on cross ROxyz (within the Registration Domain) specific functions. An example could be yearly EU reports on water quality, bridging the registration objects: groundwater-monitoring wells, monitoring networks and quality surveys. The (4) Proprietary services is a proprietary interface that connects to selected, dedicated applications. The (5) Registers support the processes of the BRO. There is one exception, the subsurface register, which reflects the prime purpose of the BRO: storage of ROxyz data. Other examples are the Transaction Register and its associated Source Document Register. Its main purpose is logging each transaction, storing the transaction data and its functional content (the Source Document). The Organisation and Mandates register keeps track of relations in the BRO: who is appointed as data supplier (delivery responsible party) for which Registration Object for which data owner (Delivery accountable party). Furthermore, the BRO stores the organisations that produce the data itself. This data producer is a container concept and entails amongst others, land surveying enterprises, maintenance organizations, owners of infrastructure, etc. The Review Register is responsible for keeping track of ROxyz are suspected of containing errors. Only one implemented (6) supporting service is the Geometry Service. It keeps track of basic geometries (like the formal borders of the Netherlands and its Exclusive Economic Zone), coordinate transformations and other geometry manipulations. There are two main (7) Administration applications. The Register Management Application is the tool the Register Administrator uses to perform his/her job. It gives access to Registration Objects, failed and aborted transactions to support data suppliers in interconnectivity, placing registration objects under review and removing them from the registration. One of the most important functions is managing the correction flow. Corrections are by nature an indicator that something is not working as it should in the interaction with data suppliers. The flow is interrupted for human intervention and the Registration Administrator had to use his/her knowledge of ROxyz to



ascertain if a correction is justified or not. Corrections are for this purpose compared side by side with the initial registered object. The Registration Administrator can accept / reject a correction and advise the data supplier in the process. The last component (8) Role Based Access Control (RBAC) has to establish the identity of the user (authentication). When the user is 'known' in the BRO and the provided certificate is valid then the rights of the user are checked (authorization). Only when the identified user has the appropriate rights for the ROxyz, the user is allowed to continue

### **BRO: 3 Layer Architecture**

The BRO adheres to a classical 3-layer architecture pattern (see [Figure 7](#)). The Access / Presentation Layer is responsible for terminating the access technology. For a web program, this layer services HTML pages. In the case of the BRO, it terminates the SOAP interface and maps the access- or product information model to the business- or domain information model. This layer checks if the received XML information is matching the schema definition (XSD). The purpose is to create flexibility, allowing different technologies to exchange data in the near future (e.g. exchange SOAP for REST). The Business Layer performs all (domain specific) business rule validations defined in the ROxyz catalogue. It is intended to span a superset of (minor) versions of a ROxyz. This layer is devoid of database or access technology and solely focuses on the business to be performed. The Data Layer maps database technology (Oracle) to a so-called Object Relational Model (ORM) and is exposed via a Data AccessObject (DAO). Its purpose is to abstract from a physical database model and implementation and leave open the possibility to exchange it for other technology. The DAO also links several related ROxyz in a domain to each other. For instance, in the groundwater-monitoring domain, it relates ground-water-levels to monitoring-networks to groundwater monitoring wells. The technology stack is based on open source technology.

The technology stack is based on open source technology. We use JBOSS Application Server that packages a set of standard frameworks, such as Hibernate (Object to Relational mapping), CXF (SOAP) and Hibernate Validation (JSR303). It is responsible for handling transaction management. The server is used as much as is possible out-of-the-box. To allow portability, whenever possible, standard J2EE and JPA interfaces are used rather than application server specific solutions.

These three layers are decoupled and have their own models. Mapping these models is done by code generation. This takes away tedious mapping work and the risk that validation rules are accidentally implemented as part of the mapping. The code generation produces readable code and is transparent in contrast to a more classic approach, which often use runtime reflection.

### **BRO Software Development**

Software production now is done using the Agile Scrum framework in a DTAP (Development, Test, Acceptance, Production) environment (see [Figure 8](#)). Once the standardization process reaches a certain maturity level, the documentation (semantic and technical data model, including definitions of operations) enters the backlog of the development team. They deliver in a steady cadence features of a web service. Slicing the documentation is done based on web service operations and a 'growing' set of business validation rules over subsequent sprints. Because full functionality of a new RO takes several sprints to complete the sprints, typically only use the dev and test environment. Continuous Build, Continuous Deployment and Continuous (Automated) Test are used to lessen the workload. The development process uses best practices (like

static code analysis, unit test, automated function test, automated build, reviewing) to guarantee and maintain software quality. The growing ROxyz will be collected in the Acceptance environment. Only when it is complete, it will be delivered to Production. This step is dominated by legislation and will only be taken when the secondary legislation comes into effect.

### **The Near Future**

The BRO system is planned to be fully implemented in 2022. Over the last eight years, the blueprint of the system has been developed and the processes of standardisation and software development have been improved. This will make the implementation of new RO's more efficient. The BRO system will not only improve the access to reliable sub surface information. Because of the growth of the number of subsurface data points (estimations expect that the number of CPT and BHR data will almost double in 5 years) the quality of the geological models that are derived from these data will improve. This will help expert users to understand the spatial variations in structure and properties of the subsurface, giving them the possibility to incorporate this uncertainty quantitatively into their own designs and calculations. Furthermore, the system will open up a whole new world of 3D exploration of the subsurface. With the web service technology and the position of the BRO within the e-government system, accurate and trusted information can be combined in real time on all kind of (mobile) devices.

Since the web services are publicly available, commercial companies are stimulated to create added value products on top of the data in the BRO, opening up a whole new market. This concept of 'open data' is a strong driver for the Dutch government to invest in the BRO. First examples already exist in a pilot stage where data from different formal registers is combined with BRO data and 3D CAD models for road traces. The user interface enables the user to explore the 3D space, viewing the infrastructure itself, the houses surrounding it and the subsurface bearing it. The last is visualised using the 3D geological voxel model, together with CPT and borehole data from the BRO. Furthermore, 3D subsurface exploration is offered to children in elementary school as 'EarthCraft', the subsurface version of the popular game Minecraft. Is this the first step to augmented subsurface reality?

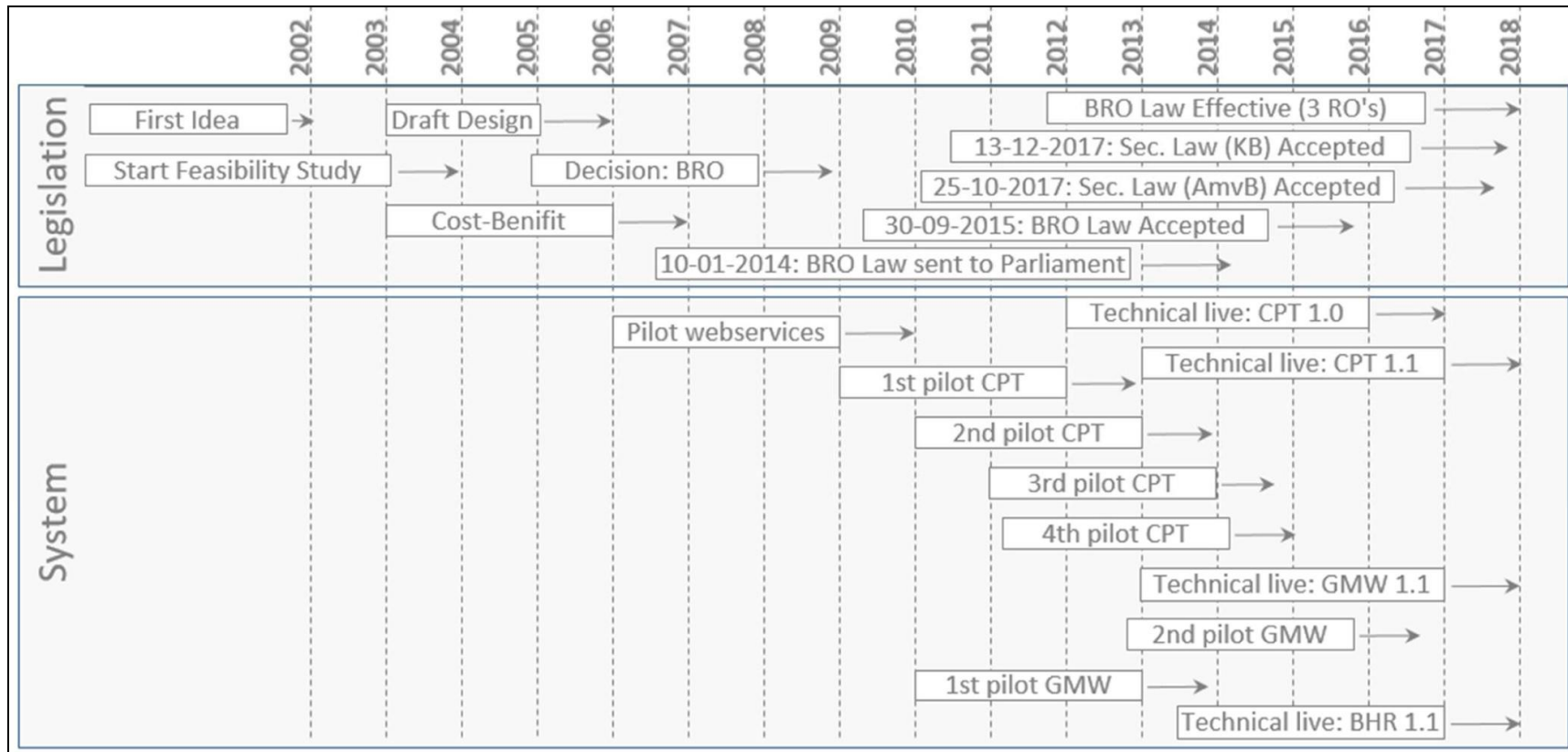


Figure 1. Timeline for the development of the BRO law and the technical system.



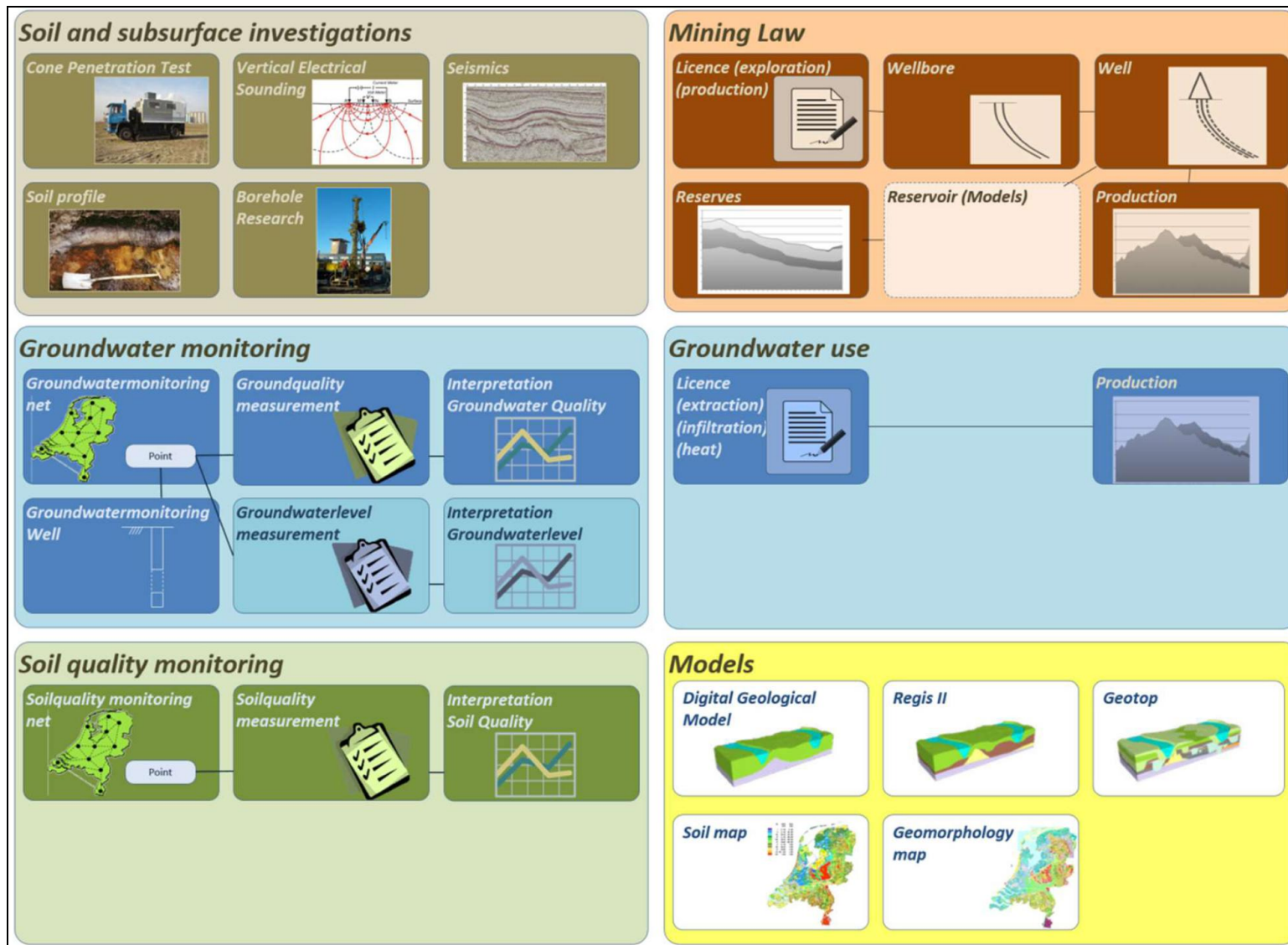


Figure 2. Overview of data domains and included RO's in the BRO.

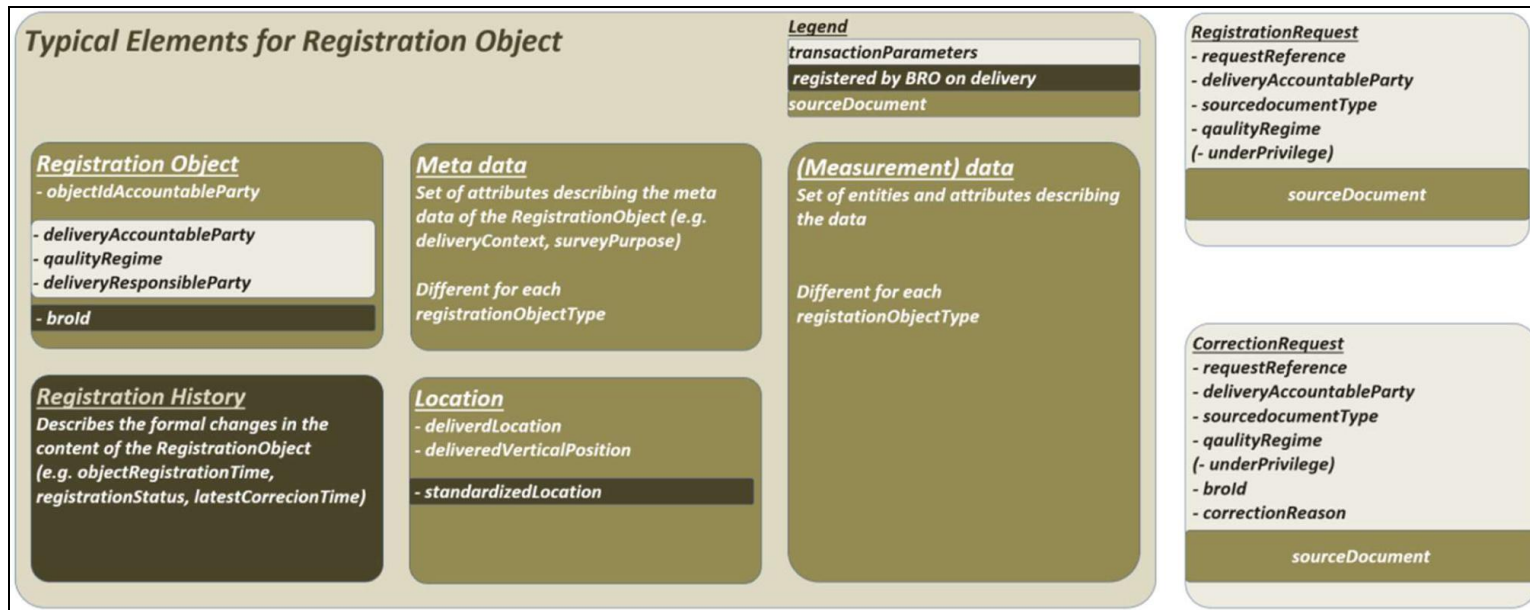


Figure 3. Overview of general elements per Registration Object

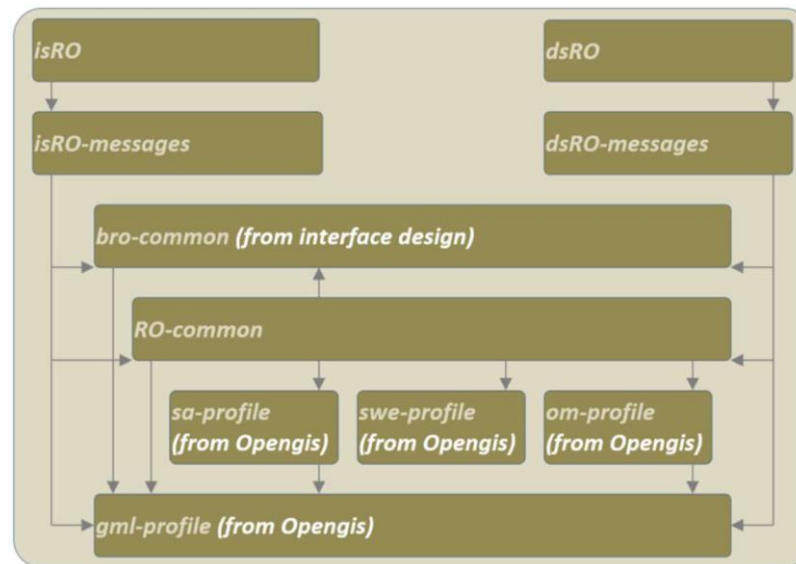


Figure 4. Packaging for a BRO Registration Object RO

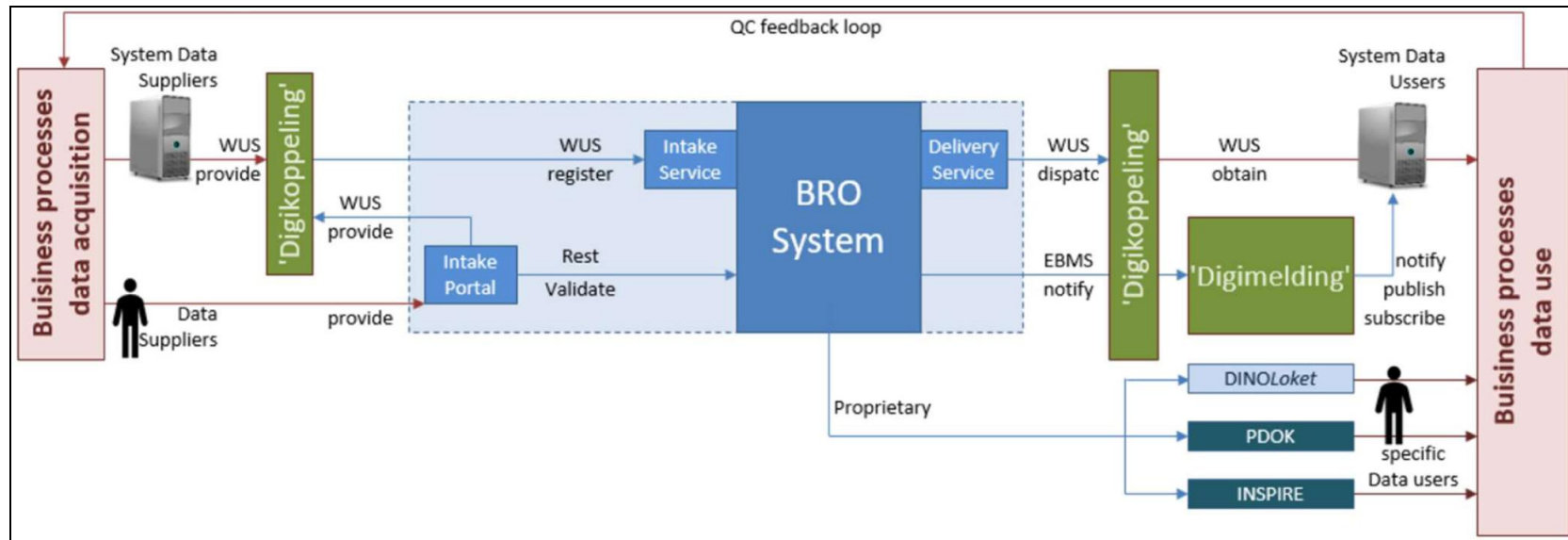


Figure 5. Interconnectivity of the BRO system (in blue), the connection to the Dutch Government network (in green), proprietary outlet portals (dark blue) and external business processes (in red).

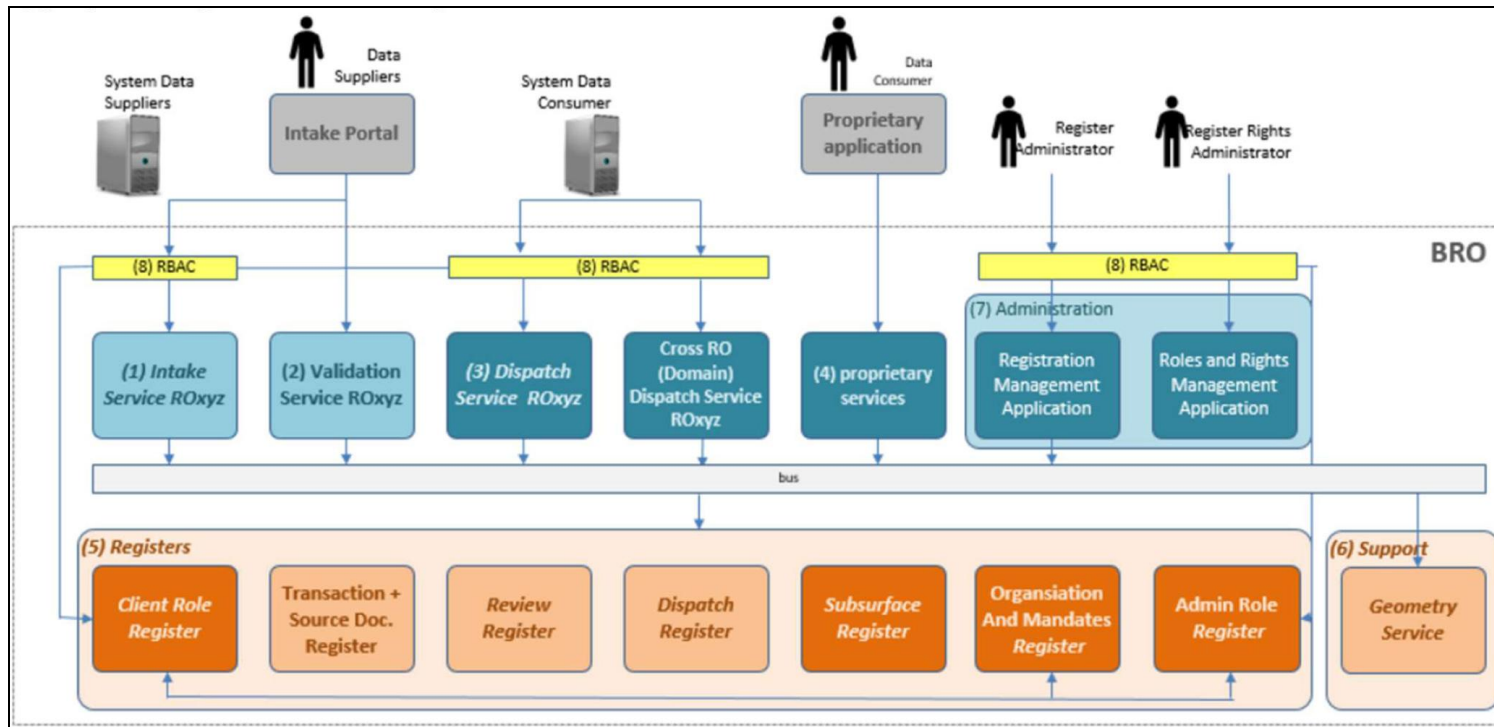


Figure 6. Functional decomposition of the BRO

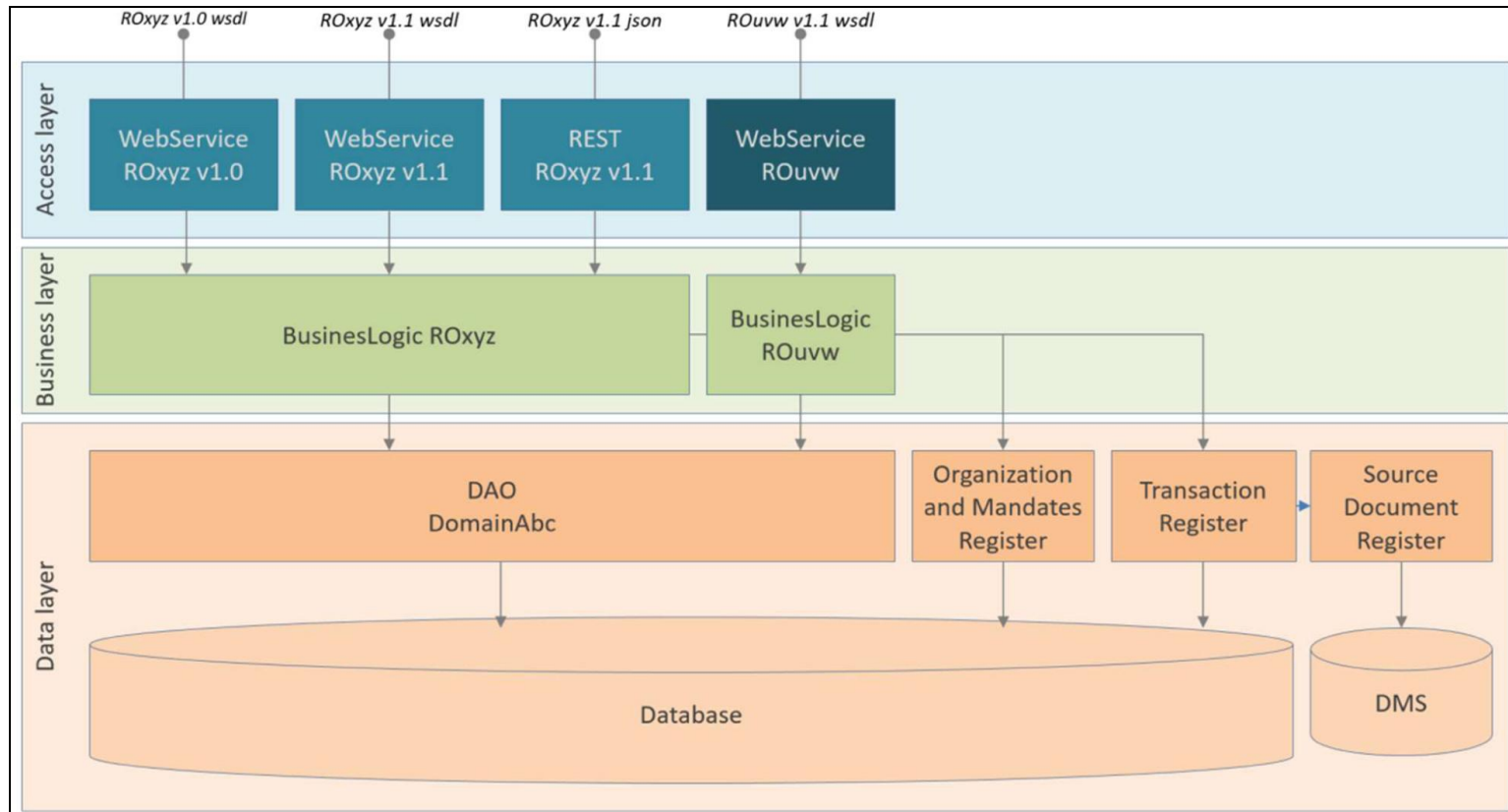


Figure 7. Three layer architecture for the intake services. Dispatch services and management applications use the same principle.

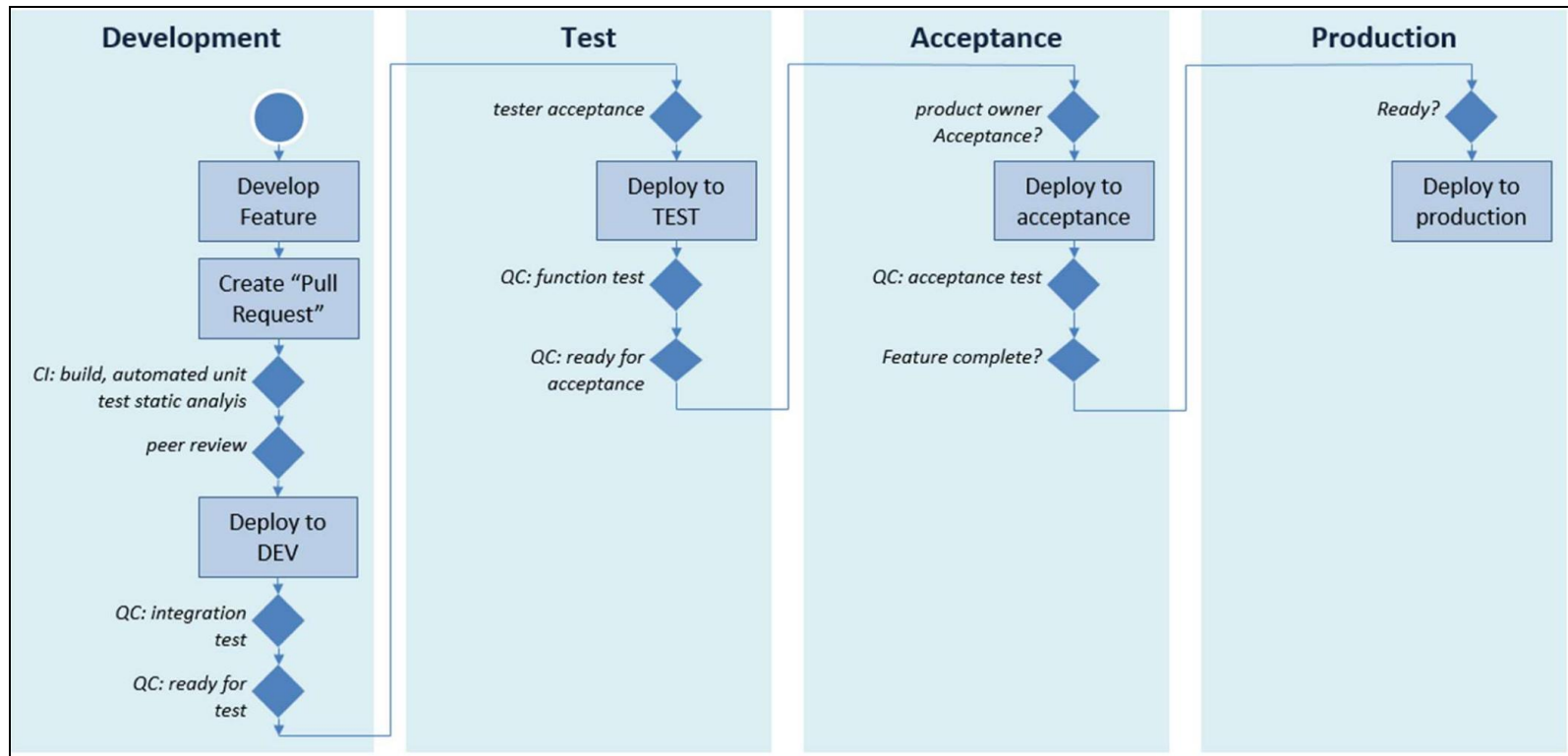


Figure 8. BRO development process.