ORIGINAL ARTICLE

Open Access

CrossMark

Geospatial binding for transdisciplinary research in crop science: the GRASPgfs initiative

Didier G. Leibovici[1*], Suchith Anand[1,6], Roberto Santos[1*] , Sean Mayes[2,5], Rumiana Ray[2], Masoud Al-Azri[2], Abdul Baten[3], Graham King[3], Asha S. Karunaratne[4,5], Sayed Azam-Ali[5] and Mike J. Jackson[1]

Abstract

The paper retraces the GRASPgfs endeavor (Geospatial Resource for Agricultural Species and Pests with integrated workflow modelling to support Global Food Security) between multiple disciplines around a common objective of facilitating research and model simulations for sustainable food security. Within this endeavor, the geospatial media has been the enabler for multidisciplinary research in crop modelling. Geospatial genetic-trait variations and associations with environmental forecasting were the main focus of the GRASPgfs. Designing the platform achieving this objective generated a transdisciplinary vision of modelling and forecasting for sustainable agriculture. Based on interoperability principles, seamless access as well as sharing for data, metadata and processing models, the design is described in this paper. This geospatial binding facilitates and supports new types of hypotheses and analysis as illustrated in the paper with a landscape genetic case study (bambara groundnut) and a crop disease modelling (eyespot disease). The approach and the eGRASP platform are generic enough to accommodate further complexity into the integrated modelling that this geospatial binding enables.

Keywords: Agriculture models, Crop modelling, Scientific workflow, Geospatial information, Genetic-trait, Interoperability, GIS, OGC, Web services

Introduction

A FAO report of the Commission on Genetic Resources for Food and Agriculture (CGRFA 13/11) clearly identified "spatial analysis to identify varieties likely to have climate-adapted traits as an aid to plant breeding" as one of the eight priorities in multidisciplinary research. In addition, Earth Observation data (EO) has proven the capacity to provide measurements of key environmental conditions to predict the production of the healthy crops and potential disease threats.

Such agricultural modelling and simulations need access to elaborated geolocated genetic-trait information as well as complementary data sources coming from geospatial data providers and geospatial data hubs, e.g. soil moisture data, climate data. The BBSRC funded project GRASPgfs "*Geospatial Resource for Agricultural*

Species, pests and Pathogens with workflow integrated modelling to support Global Food Security" initiated the design and prototype of an e-infrastructure linking together (i) a geo-germplasm database, (ii) a dynamic metadata catalog and (iii) a workflow modelling tool to enable optimal usage of the geo-genetic-trait information. This is desired in various types of simulations for crop trait variation, forecasts under climate forcing scenarios and crop disease epidemics.

This initiative and the prototype of the e-infrastructure based on open standards is described.

Background

Access to a wide range of information, from rigorous scientific results to 'hear-say' farmers' knowledge is becoming critical to be able to target efforts in food security planning at community or at country levels. Also, designing scientific and intervention strategies within changing climates and markets is a fundamental

* Correspondence: Didier.Leibovici@nottingham.ac.uk;
Roberto.Santos@nottingham.ac.uk
[1]Nottingham Geospatial Institute, University of Nottingham, Nottingham, UK
Full list of author information is available at the end of the article

Leibovici *et al. Open Geospatial Data, Software and Standards* (2017) 2:20

Page 2 of 11

challenge. The Plant Genetic Resources for Food and Agriculture part of the FAO's Global Plan of Action for plant genetic resources has been established as a worldwide dynamic mechanism using WIEWS[1] to foster information exchange among members (more than 150 countries) and as an instrument for periodic assessment.[2] The enhanced use of this type of resource and other germplasm banks with additional contextual information is nonetheless still highly desired. In a recent FAO report[3] from the Commission on Genetic Resources for Food and Agriculture clearly identified *"spatial analysis to identify varieties likely to have climate-adapted traits as an aid to plant breeding"* as one of the eight priorities in multidisciplinary research.

Technologies for the collection and dissemination of geolocated information, using broad-band mobile communications, sensor platforms, spatial search and pervasive computing, are fundamentally changing the access to and use of location-based data in agriculture [20]. However, the necessary cross-disciplinary research needed to transform raw data and information into useful intelligence and knowledge to improve the planet's environmental, economic and societal well-being is still constrained by disciplinary and organizational silos and legacy concepts. Even if this was already acknowledged in the 1940's, the geo-location of genetic data in ecology and agriculture for further spatial analysis: spatial genetic, is still a recent concept [17]. Within the *"from farm to fork"* chain, various heterogeneous data including genetic-trait information are to be considered as part of the computational modelling for prevision and forecast; most of them have a geo-location or spatial component or would be required to have one to be used from a plethora of model applications of various complexities. These are either biophysical, agro-economically based and more mechanistic or deterministic orientated [6, 13, 14, 19, 30, 36] or more stochastic orientated [2, 5, 12, 23, 42, 48, 52], more rule-based including agent-based orientated [32, 45, 49] but all contain a combination of those types. Therefore, a cross-disciplinary expertise driven from geospatial sciences methodologies appeared to be needed to develop an integrating framework for relevant data sources, in order to allow knowledge gathering across all subjects relevant to Food Security.

The objective of the geospatial integration e-infrastructure framework that GRASPgfs initiative has proposed to establish is to facilitate the use and reuse of trait data in crop, animal and microbial species of agricultural importance. The initiative relies on the position that the geospatial realm, as an entry point and end-point of this e-infrastructure, enables researchers but also stakeholders and policy makers to ground their development strategies but also to elaborate more easily alternatives [3, 6, 26, 30, 31, 36, 40]. Not only this binding

is conceptual and interdisciplinary but also it has conrete technical impacts on the e-infrastructure utilizing Open Geosptial Consortium (OGC) standards. Linked with dynamic climate records within the framework capabilities on scientific workflow modelling, this would allow addressing food security issues for sustainable agriculture by enabling predictive modelling with identification and characterization of new sources of germplasm.
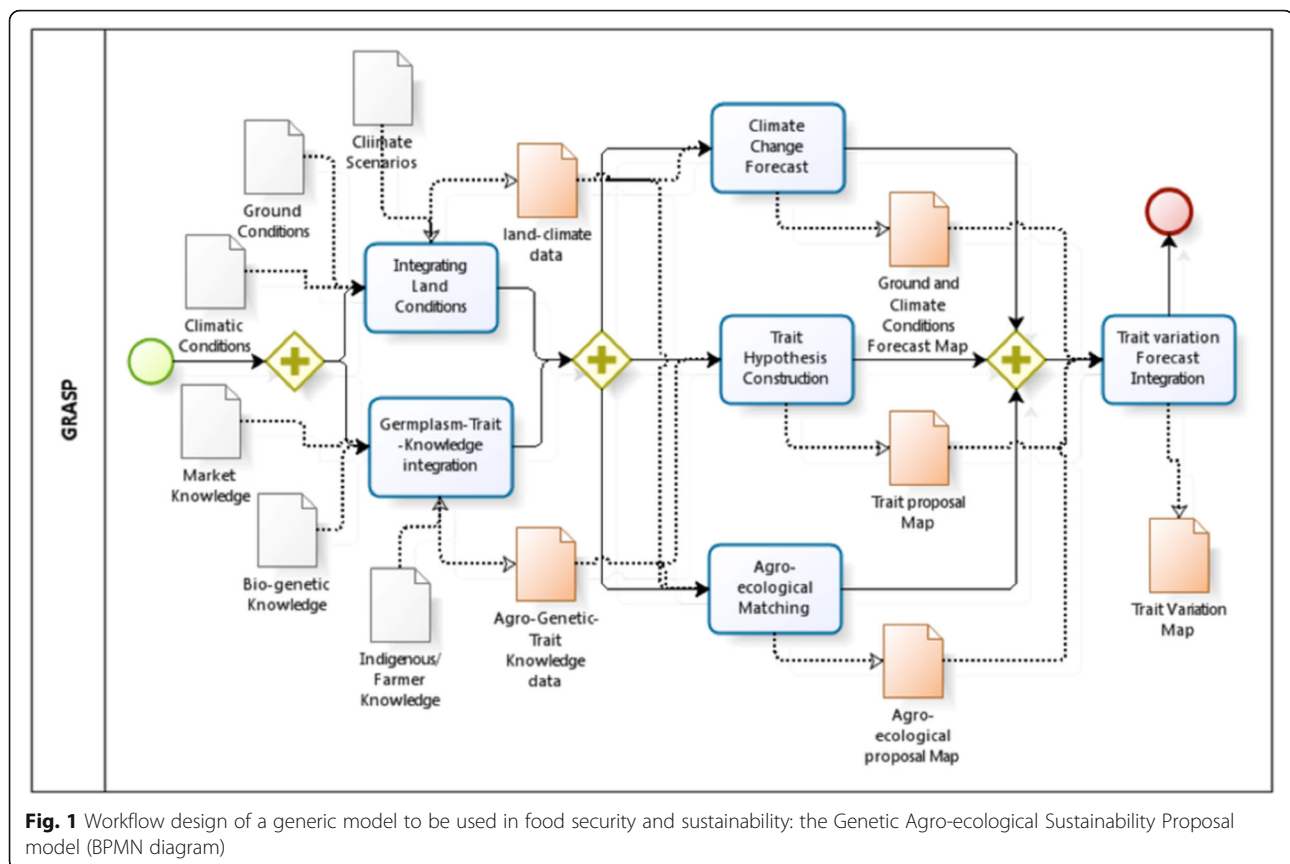
The paper describes the initial overall architecture design and first results on establishing the *eGRASP* platform and e-infrastructure. Section 2 concentrates on expressing the method in defining the approach and initial challenges for long-term objectives; section 3 highlights the bases in designing the *eGRASP* solution and architecture along with illustrative examples of initial results on using such approach.

Requirements

Targeting global food security issues and sustainable agriculture, related to crop selection and climate change needs the development of models integrating a range of disciplines such as genetic, agro-ecological modelling and land-climate forecasts. Geospatial science can be the mediating component of an *e*-infrastructure enabling data and processing to be retrieved, integrated and made available within a geospatial scientific workflow modelling interface with uncertainty management.

The main objective of the geospatial integration framework wihin GRASPgfs was to facilitate the use and reuse of known (and new) sources of crop traits together with dynamic climate records within the framework capabilities of workflow modelling addressing food security issues concerning sustainable agriculture. Describing agricultural species germplasm for genotype characteristics with the data ordered by geospatial origin, the higher-level descriptor being "agricultural trait" has been put forward as enabling new way of expressing and analysing trait variations [34, 41].

Highlighting model complexity for integrated assessments required for global food security, a recent review of crop models under climate forcing pointed out the need of generic solution enabling or facilitating the combination of various models together [14]. Figure 1 encapsulates the challenges of facilitating the elaboration of such analysis via an integrated workflow modelling. This workflow grasping the "big picture" and illustrating the cross-disciplinary expertise required is conceptual and each data entry or processing task may in fact illustrate the use of complex data structures and sub-workflows themselves. The framework to develop would need to facilitate the integration of the driving key conceptual aspects of this model. Following this conceptual approach and being able to re-use data and models available to instantiate such model would enable new perspectives on

Fig. 1 Workflow design of a generic model to be used in food security and sustainability: the Genetic Agro-ecological Sustainability Proposal model (BPMN diagram)

crop genetic diversity by (i) identifying new sources of trait variation, (ii) geolocating suitable germplasm, (iii) planning breeding objectives with the greatest likely impact from the added information of local market and farmer's knowledge [8, 34], and (iv) evaluating the effects of climate change scenario.

Integrating heterogeneous datasets coming from various sources within a generic platform means being able to access and understand the semantics of these data and processes in order to allow the platform to present the data, analyse them, or instantiate a workflow model using them [9, 25]. Easily integrating various data and processing resources has considerable advantages in terms of rapid development of models and their execution but gives less control on the quality of the results as various uncertainties coexisting in the components of the workflow model. Therefore, bounding with uncertainty assessment the outcomes of the models should also be the aim of the *eGRASP* platform to allow better decision-making. Specific capacities are also needed to integrate information such as genetic-trait encoding and ontology binding with disparate germplasm data sources. Pests, pathogens and weeds are encompassed at the crop information level and in the process themselves as they are often impacting from interacting with the environmental conditions.

Designing, developing and implementing

The GRASPgfs has therefore focused on designing and implementing a flexible, interoperable platform based on open source software[4] compliant with GEOSS[5] using OGC[6] standards and services for data and processing capabilities. From delivering a flexible, integrative and sharing *eGRASP* web platform based on openess, the objectives of enabling researchers in crop modelling, agro-ecological modelling either as developer of new models or evaluating agriculture strategies (agro-ecomic modelling), to seamlessly re-use existing models and specific data such genetic-trait information will be achieved. For efficiency and controls on the quality in terms of uncertainty and variability of the outcomes, the design of the platform allowed functionalities to easily browse and visualise metadata as well as has to geo-computationally evaluate workflows output uncertainties [15, 25, 28]. Spatial analysis of the spatial variations either of the predicted outcomes and their uncertainties were included in the design to be part of the platform as well. That way the modelling part and of the decision making part are interlinked, allowing more flexibility and adaptability. The approach and the concept of the *eGRASP* platform has been the result of multidisciplinary exchanges leading to a real transdisciplinary vision [4, 21, 38] that is highlighted in the next section.

Leibovici *et al. Open Geospatial Data, Software and Standards* (2017) 2:20

Page 4 of 11

Emergence of a transdisciplinary vision

Whilst building up a core collaboration on this topic from a range of disciplines (within environmental and human geography, crop science, geospatial information, and computing science) at the University of Nottingham by meeting regularly and having small funding for a few summer internships in 2010, the common vision expressed in Fig. 1 started to emerge. Later on, thanks to a 18 months pump prime funding from the BBSRC the research work could start. The workflow of Fig. 1 encapsulates the vision put into the design of the *eGRASP* platform as much as it is a template of potential modelling scenarios envisioning the various components as data and processes needed to consider fulfilling our objectives for GRASPgfs. If at first it may have seemed that the geospatial sciences brought tools enabling this research within a cross-disciplinary perspective, it transformed rapidly into acting as a media of a more holistic integrated approach [16], which then expressed itself in challenging its specific developments within a context beyond the disciplines involved. In addition to providing more opportunities for expanding the capabilities and applications looked for in the first place, this advancement also created new avenues for interdisciplinary research and practices in the use of GIS in agriculture research.

Beyond the global concept and concepts encapsulated in it, Fig. 1 is a truly transverse vision that not only put each specialist of a sub-model within a contextual flow but also enriches the geospatial e-infrastructure modelling framework. This resulted from various flow diagrams of conceptual information into a technical and standardised representation using a cross-disciplinary encoding standard, the BPMN standard (Business Process Modelling Notation from the OMG standard organisation). As far as the cross-disciplinary concerns, Fig. 1 as a BPMN representation is also a scientific geo-computational model seen from a meta-level description that can be linked to a workflow engine enabling its computational execution once instantiated (Fig. 2).

In order to instantiate such models (entire Fig. 1 or sub-models encapsulated) the design of the eGRASP platform is based on the Use Case model in Fig. 2, which translates the requirements exposed earlier. In this figure only general use cases are presented with different colours to express the different domains or disciplines
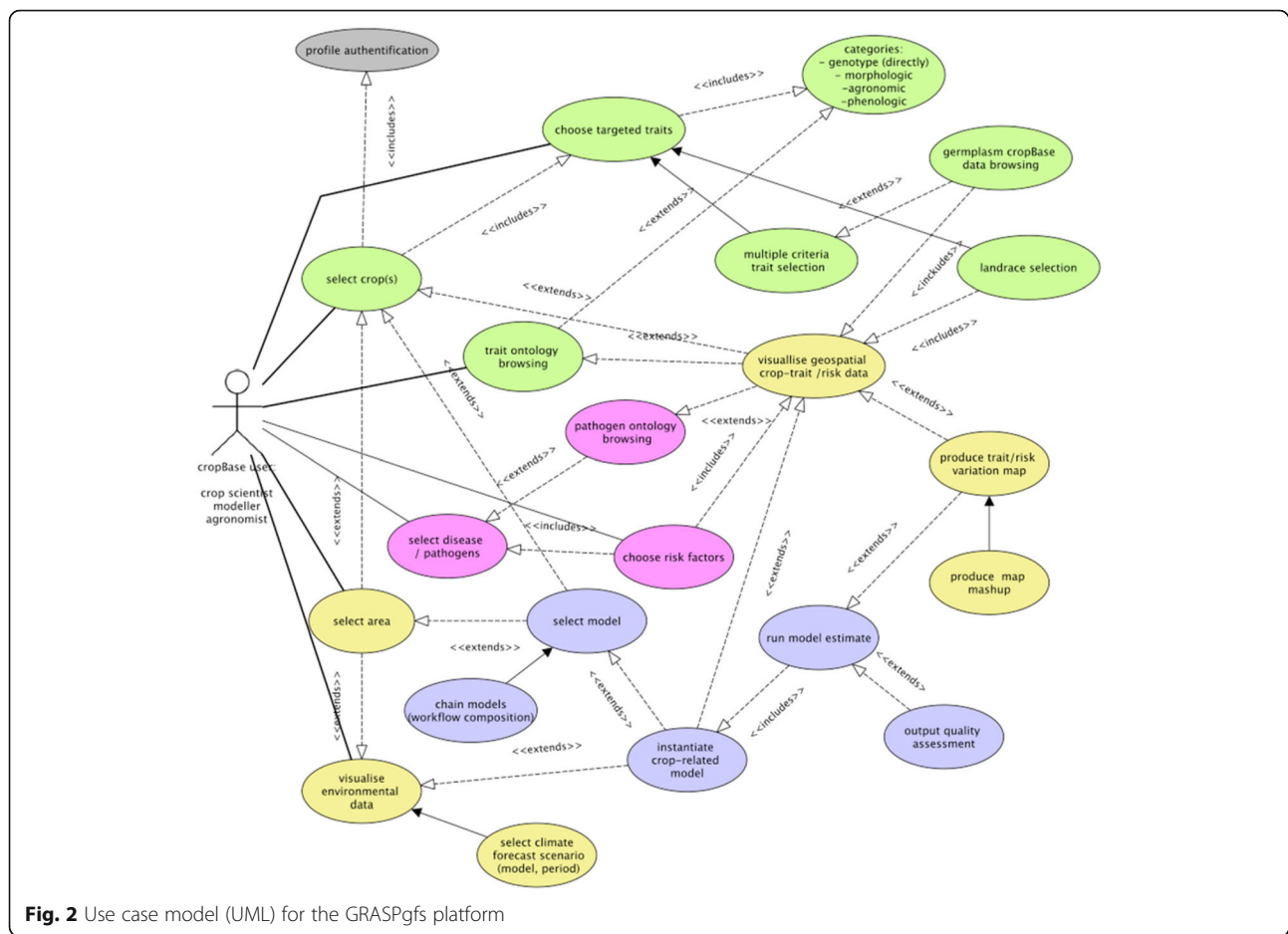

Fig. 2 Use case model (UML) for the GRASPgfs platform

Leibovici *et al. Open Geospatial Data, Software and Standards* (2017) 2:20

Page 5 of 11

concerned: the green use cases reflect the crop genetic with genetic-trait information aspects, the yellow use cases concern geospatial science with visualisation and selection of environmental constraints, the bleu use cases are to do with geocomputational modelling and scientific workflow composition and evaluation, and the pink use cases concerns crop epidemiology with the risk factors associate with the crop modelling including pests and disease risks from pathogens information.

Like UML[7] (Unified Modelling Language) particularly using class diagrams for object modelling and use case diagrams such as in Fig. 2, has been enabling cross-disciplinary exchanges from data modelling [22], the BPMN language establishes a bridge between the conceptual integrated modelling towards the effective execution of the models [44]. Facilitating the composition of such workflows using existing resources is paramount [11].

Crop modelling complexity

Well-known crop modelling approaches such as APSIM[8] [19], AquaCrop[9] are considered here as expressing or being a sub|-model of the *"trait variation forecast integration"*. The purpose of the GRASPgfs is to re-use

directly these established models within a flexible platform; they can be wrapped into OGC web processing services (WPS) and made available for the platform as such [10, 35] or via a brokering system [7, 39]. When the models can be broken down into sub-components, if required by the crop-trait variation scenario, this can be made available to the processing service. When possible the interaction of these models can be complex to set up and to combine, the BPMN editor is seen as a simplification, particularly when a few models are to be combined. Ultimately it brings interoperability in interfacing heterogeneous data and processing models that do not necessarily impose standardisation for each of them. This does not preclude of course a good understanding of the models used, but the goal of the *eGRASP* platform is to hide this complexity and to focus on the ability to re-use the resources within a more macro scenario for global food security. Models and types of models identified in introduction can be potentially re-used here and the platform objectives are also to facilitate their encapsulation as WPS services (Fig. 3).

When looking at trait variation with genotypic information, the crop modelling may start with building up a selection for trait-variation linked to genotype linkage
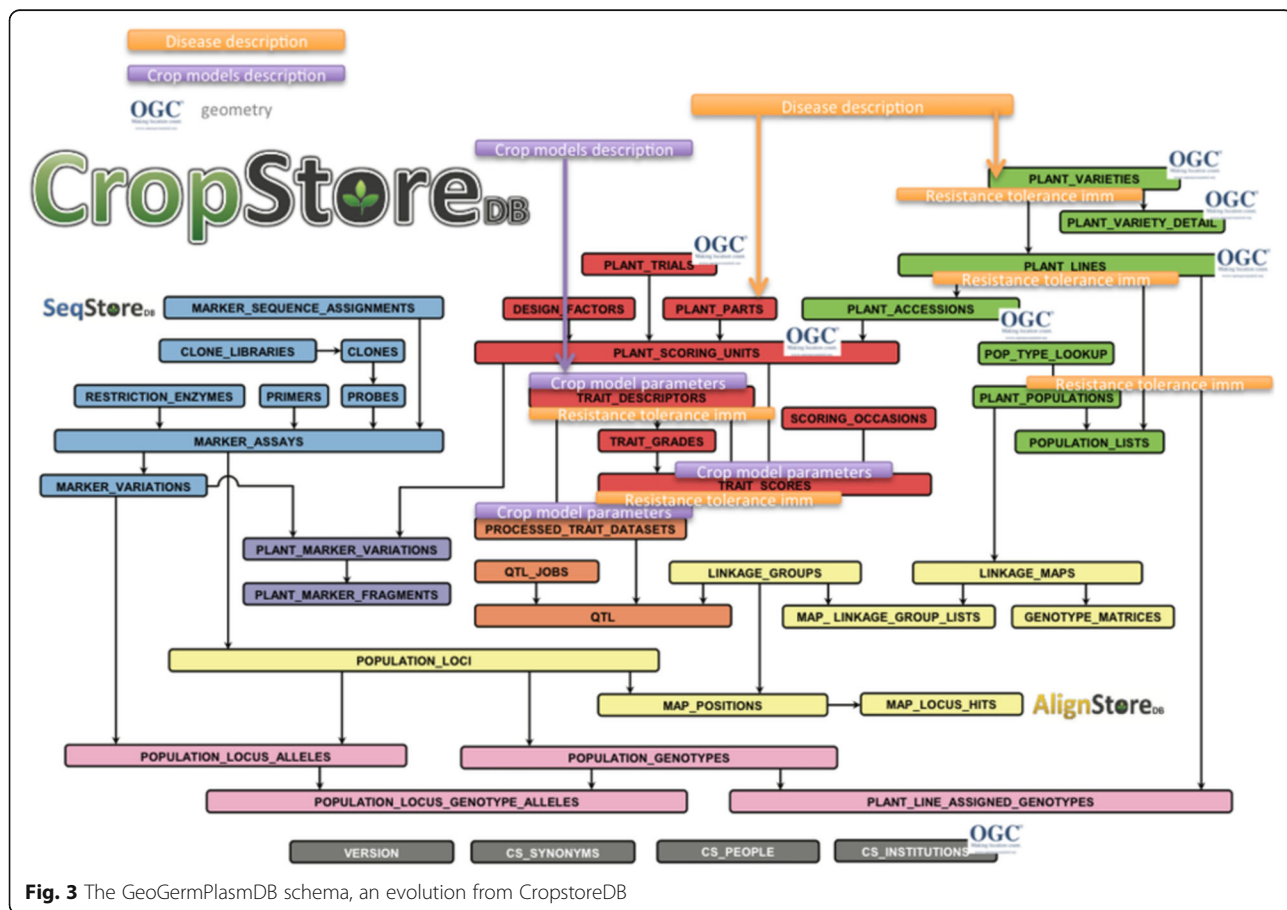

Fig. 3 The GeoGermPlasmDB schema, an evolution from CropstoreDB

Leibovici *et al. Open Geospatial Data, Software and Standards* (2017) 2:20

Page 6 of 11

and environment interaction. This corresponds to the *"Trait Hypothesis Construction"* process task in the generic workflow. To this end, it is described in Fig. 2 among the other capabilities of the *eGRASP* platform; the functionalities associated to this genetic-trait selection, before performing the crop modeling for example, are the green part of the use case model. To achieve this aspect the platform is reusing the CropStoreDB[10] database, called GeoGermplasmDB in the architecture design (Fig. 4). The GeoGermplasmDB has an extended schema in order to record the geometry associated to few tables using the OGC standard (Fig. 3) and also to be able to encode the pest and pathogens characteristics along with model parameters associated to the crop varieties as stipulated in the requirements. The Geogermplasm DB allows users to record genotype information and trait information with geo-location depending on the origins of the seeds and the trial sites and implements the component *"Bio-genetic Knowledge"* component of the platform. Geospatial variations associated with genetic variations can lead to breed selection

programs [18, 33]. An example using the underutilized crop of the Bambara groundnut (*Vigna subterranea*) is detailed in the example section (Fig. 5).

The other aspects of complexity considered here come on one hand from the interaction of farmer's knowledge with respect to the land races linked to their strategies to make a living [24, 32] and on the other hand to the climate forcing interacting with the current land conditions. Due mostly to aggregation and topological properties when modelling these models, the spatial complexity can now be also introduced [26, 47, 51]. Specific models for climate forcing more often mechanistic can be used to predict future ground conditions but are usually integrated with interaction from general land use categories [43, 50].

The eGRASP capacity
The approach pursued in GRASPgfs and for the design of the *eGRASP* platform has been as much top-down as bottom-up from leading disciplines such crop genetics, geospatial information modelling and crop modelling.
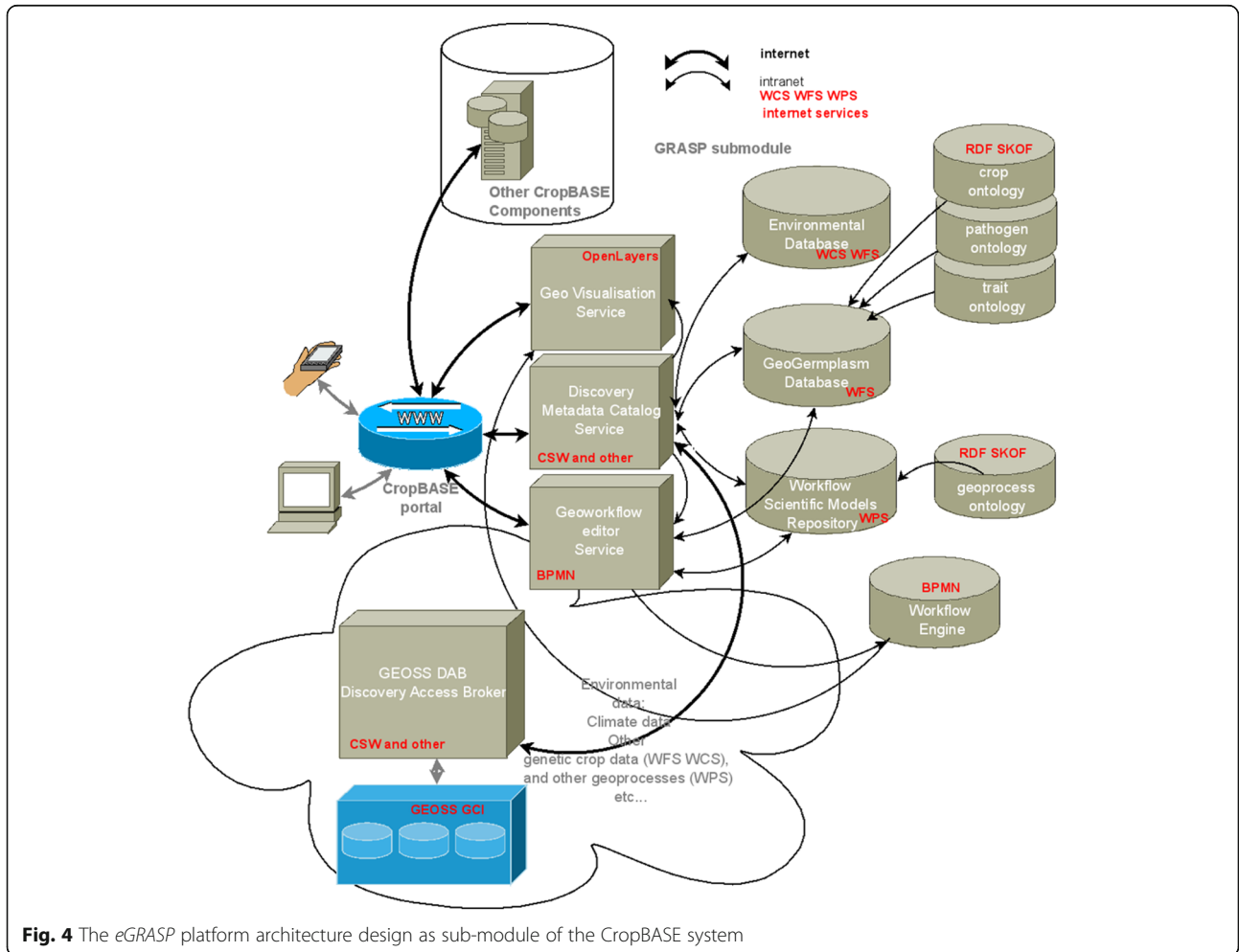

Fig. 4 The *eGRASP* platform architecture design as sub-module of the CropBASE system

Leibovici *et al. Open Geospatial Data, Software and Standards* (2017) 2:20

Page 7 of 11

Fig. 5 A landscape genetic analysis workflow on bambara groundnut (*Vigna subterranea*) landraces

Basically besides strong top-down emphasis on geolocated genetic-trait database (the GeogermplasmDB), and on a workflow modelling (based on OGC WPS and BPMN standard), case studies analysis were used to gather requirements. Mixing these two aspects as well as envisaging direct use of the top-down elements into the bottom-up approach, the UML use case diagram of the required functionalities of the *eGRASP* platform was obtained (Fig. 2). From the adoption of the use case diagram, disciplinary research took place to refine the case studies with focus on use case matching and potential new developments whilst the computing architecture was design to fit these requirements.

The architecture designed for the *eGRASP* platform to enable global spatial data infrastructure functionalities, as well as the ones described above, is given in Fig. 4. This viewpoint gives an overview of the different components without detailing on how specific analytical functionalities are implemented. The objective for this pump-prime funding was to establish the design and to demonstrate a prototype. Therefore, specific functionalities are still to be developed; further funding is required to pursue these efforts. In Fig. 4, front-end services with their clients are represented as square boxes and back-end services often associated with specific information (e.g., databases, repositories) are represented as cylindrical boxes. The *eGRASP* system appears in this design as sub-architecture of the CropBASE[11] initiative led by CFF (Crops For the Future), a wiki-knowledge sharing platform integrating multiple CFF programs also in development.

For the sake of demonstrating the architecture the set of services implemented and facilities currently available,[12] but the platform as well as the CropBASE portal are not yet operational. The OGC services, for example using WPS and WFS, can also be used directly in other clients such as in QGIS (from the OSGeo[13] stack), currently:

- the Geovisualisation is supported from QGIS and from the WMS client provided from the Geoserver serving the GeogermplasmDB
- the Discovery via Metadata Catalogue service (OGC CSW) is supported by GeoNetwork[12], queries on GEOSS registered catalogue can brings re-usable resources (data or processing services) as well a s local ones.
- the GeoWorkflow is supported by a bespoke specification for OGC services using the jBPM[14] suite with a web editor and a workflow engine [35].
- the GeoGermplasmDB services as well as local environmental data are served using GeoServer[12]; the results of the simulations or other workflows can be stored in the local environmental data storage.
- a set of ontologies can be used to enrich the data and processes enabling refined queries via the metadata catalogue client.

Quality information available for data and processes in the metadata catalogue are used for uncertainty

Leibovici *et al. Open Geospatial Data, Software and Standards* (2017) 2:20

Page 8 of 11

assessmet from the error propagation, by then allowing better decision-making. This is currently available as added functionality from the web editor from re-using the MetaPUnT WPS[15] service [27, 28] and allowing to meta-propagate the uncertainties.

First applications

Two illustrative examples are presented here to highlight the potential of the *eGRASP*. The first one, a landscape gentic modelling, uses directly the GeoGermplasmDB and WFS associated to describe spatially genetic distances of germplasms. The second one illustrates the crop disease modelling of the *eGRASP* facility by designing an examplar wheat eyespot disease model [1]. Both examples, the landscape genetic association analysis and the crop disease modelling are using a BPMN scientific workflow representation, by then demonstrating the range of modelling situations that *eGRASP* is intending to cover.

For the landscape genetic modelling, a glasshouse trial with 128 plants from 4 repetitions each of 32 landraces was analysed (Figs. 5 and 6). Here only the genotypic information was used to retrace geo-location associations of similar genetic profiles based on 20 microsatellites molecular markers (SSR) [37, 46]. Five genetic profiles were identified from k-means on main principal components of the SSR response data. In Fig. 6, the green and red profiles capturing most of the genetic variability are relatively clustering spatially with an East-west gradient in the Sahel for the reds and a North-south gradient in the East and South-East Africa for the greens. Adaptations to similar climatic environment can be though as

explaining these zones with the Sahel zone for the reds and a more humid tropical zone in the East-Africa for the greens. Trade routes can be also involved. Further analysis including the phonologic data with comparison to local data will be needed to confirm these sorts of hypotheses.

Each task of the workflow in Fig. 5 was performed from R scripts based on existing packages. These R scripts are in the process of being encapsulated as WPS in order to be used and shared from the *eGRASP* platform.

The second example illustrated in Fig. 7 is a scientific workflow for crop modelling with potential occurrence of the eyespot disease. The purpose was to integrate specific epidemiological disease modelling within a normal growth simulation model. The Eyespot disease is modelled using few sub-models interfering with the normal development of the crop:

- The inoculation potential model (IPM) determines the amount of inoculum available for infection of the host depending on land condition risks and weather data.
- The disease development model (DDM) based on the inoculation level and key environmental factors related toinfection and disease developement.
- Finally at a key developmental growth stage the severity of the disease is determined (DSM) and is based on estimates from the previous two models.
- The impact of the severity of disease is then evaluated iteratively (HRM) at the subsequent growth stages until the crop has been harvested.
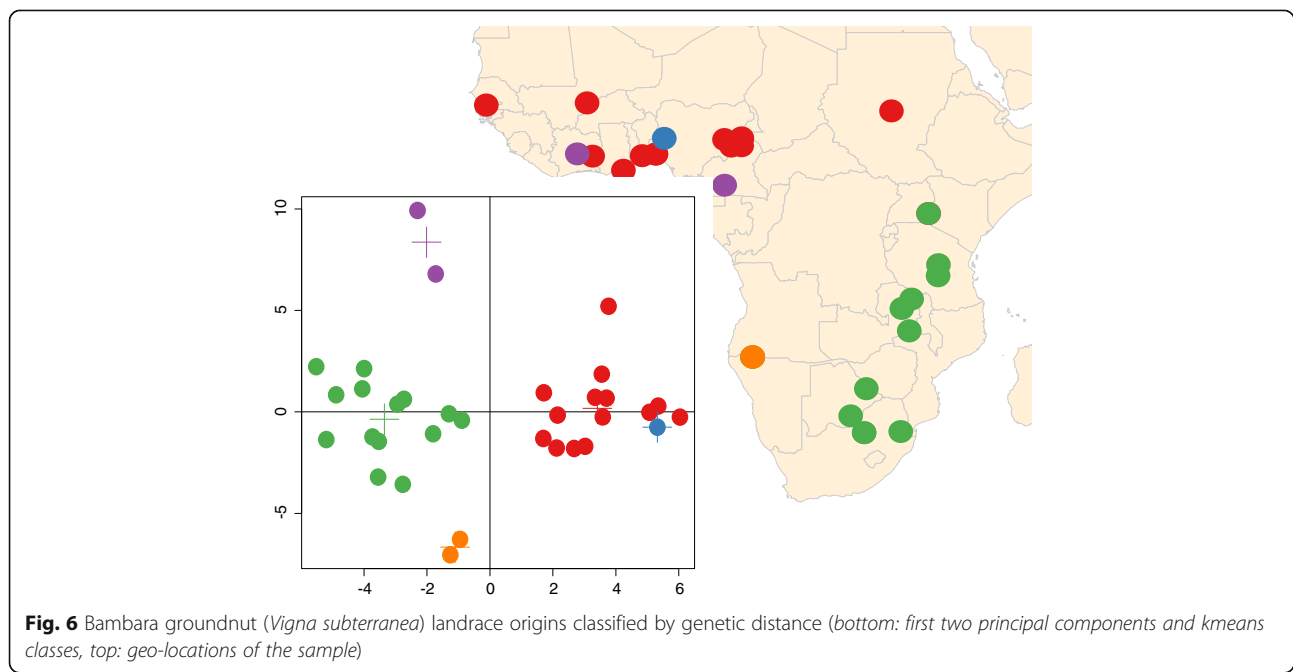

Fig. 6 Bambara groundnut (*Vigna subterranea*) landrace origins classified by genetic distance (*bottom: first two principal components and kmeans classes, top: geo-locations of the sample*)
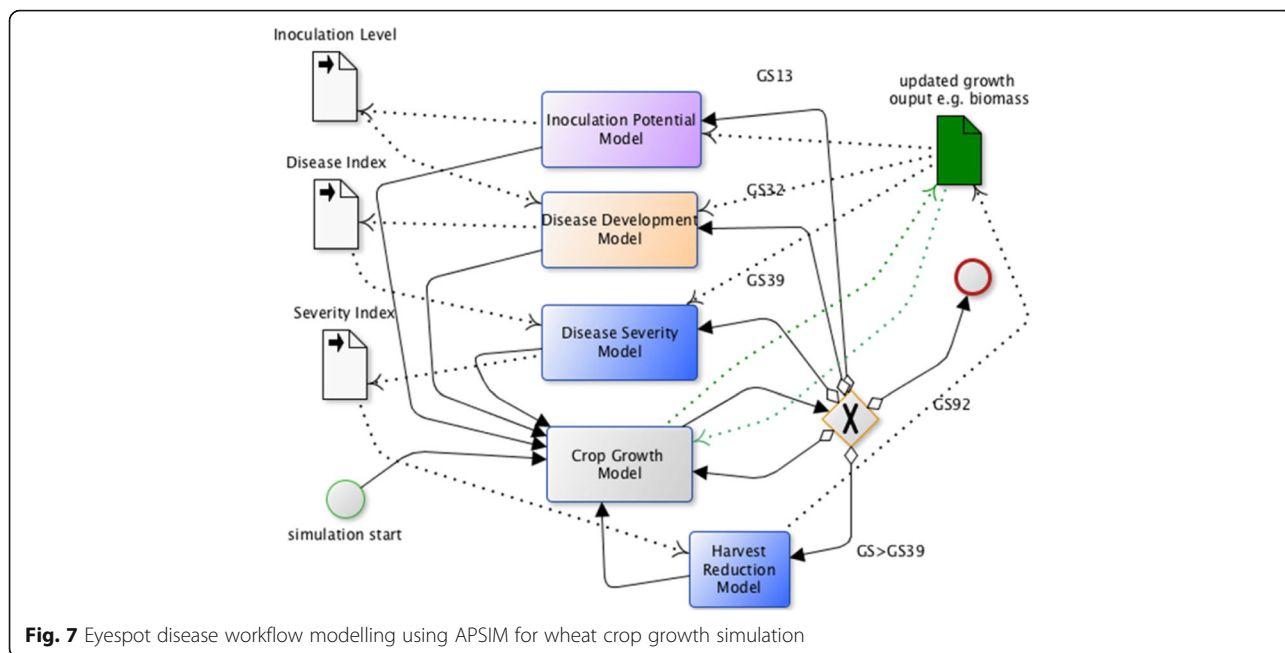
Leibovici et al. Open Geospatial Data, Software and Standards (2017) 2:20

Page 9 of 11

Fig. 7 Eyespot disease workflow modelling using APSIM for wheat crop growth simulation

Each one of the models: IPM, DDM, DSM and HRM are stochastic models and estimated at the given growth stages that were identified as crucial during the development of the crop on controlled data: GS13, GS32, GS39 and GS65 [1]. The models are to be combined with physiological based model for crop growth as in the BPMN representation in Fig. 7. The disease evolution models have been implemented in R[16] and APSIM was chosen as crop growth model. Within APSIM and using the script manager, R scripts can be ran, making APSIM the orchestrating engine. Nonehteless, encapsulating APSIM within a WPS could be a future solution using the workflow engine wihtin eGRASP. Details of first results and variables involved in the IPM, DDM, DSM and HRM model can be seen in [1] as well as the full validation of the models. Nonetheless, despite the capacity of APSIM to run R scripts, the targetted variables by the disease modelling couldn't be updated during simulations which led to a much simpler adaptation of Fig. 7.

For the eGRASP the interest lies in the fact that such composition and conceptualisation of the models can be facilitated and controlled, e.g. looking for model adequacy. The interoperability ensures that the models designed according to the BPMN standard can be then shared using a standard graphical representation for better communication but also as XML encoding enabling any workflow engine to run the scientific model represented as a workflow.

Like UML (Unified Modelling Language) used as a computing science tool to design of application systems, leading both to databases and object programing

implementations, the meta-language of the BPMN can be very rapidly understood from the scientists involved [22, 29]. This transdisciplinary process enabled to conceptualise the disease evolution and impact in a comprehensive way that has been also efficient to put in practice once each sub-model (tasks in the BPMN diagram) has been established and fitted.

Future research

Interdisciplinary projects often reduce to cross-disciplinary spill over; however over a shared building up initiative to advance on the GRASPgfs concept, a real transdisciplinary collaboration has been initiated and experienced. Not only the co-design of the eGRASP platform with its embryo of capacities has enabled to envisage new potential research ideas in each of our disciplines, it also concretised global food security strategies and analyses. The recent development of the GRASPgfs framework along with the design of the eGRASP was limited as due to the budget and not all the disciplines firstly envisaged could be adequately integrated. Whilst in Fig. 1, the agro-ecologial interaction would derives mostly from re-using models in landscape genetic and landscape ecology, as well as the agro-economic would benefit from models mentioned in the background section, their data modelling integration represented on the left hand side of the model has not been yet investigated. For a prototype design this was not crucial as long as we could still represent its future influence when composing the models.

Leibovici *et al. Open Geospatial Data, Software and Standards* (2017) 2:20

Page 10 of 11

If some of the services in Fig. 4 are in place the actual data and processes content is rather small as this was a proof of concept exercise. Nonetheless PhD students and recents projects are providing valuable examples also enhancing the capacity of this platform. The interoperability principle adopted by the *eGRASP*, including its open source and open standard focus, is the chance for maximum dissemination of this capacity as a set of cross-platform clients and services. Geospatial risk assessments in agriculture in relation to species and pests, can be greatly facilitated from sharing data and processes which can then reused by the *eGRASP*.

Endnotes

[1]http://apps3.fao.org/wiews/wiews.jsp?i_l=EN
[2]http://www.fao.org/agriculture/crops/core-themes/theme/seeds-pgr/sow/sow2/en/
[3]CGRFA-13/11/Report: http://www.fao.org/docrep/meeting/024/mc192e.pdf
[4]www.osgeo.org
[5]www.geoportal. org
[6]www.opengeospatial.org
[7]www.UML.org
[8]www.apsim.info
[9]www.fao.org/nr/water/aquacrop.html
[10]www.cropstoreDB.org
[11]http://www.nottingham.edu.my/CFFRC/Projects/CropBase.aspx
[12]https://grasp.nottingham.ac.uk
[13]www.osgeo.org
[14]www.jbpm.org
[15]geoprocessing.nottingham.ac.uk
[16]www.r-project.org

Abbreviations

APSIM: Agricultural Production Systems Simulator; BPMN: Business Process Modelling Notation; CFF: Crop For the Future; CSW: Catalog Service Web; DDM: Disease Development Model; DSM: Disease Severity Model; *eGRASP*: e-infrastructure for GRASP; FAO: Food and Agriculture Organisation; GEOSS: Group on Earth Observation System of Systems; GIS: Geographical Information Science; GRASP: Geospatial Resource for Agricultural Species, pests and Pathogens; GRASPgfs: GRASP with integrated workflow modelling to support *global food security*; GS13: Growth Stage 13; HRM: Harvest Reduction Model; IPM: Inoculation Potential Model; JBPM: JBoss Business Process Modelling suite (Java); OGC: Open Geospatial Consortium; OMG: Object Modelling Group; OSGeo: Open Source Geo; SSR: Simple sequence repeats; UML: Unifed modelling language; WPS: Web processing service; XML: eXtended Markup Language

Authors' contributions

DL and SA have been leading scientifically the project. DL finalised the design of the eGRASP and wrote the paper, both based on contributions from all. Particularly, SM, GK AB, RR, RS, AK provided work, expertise and information on the database genetic materials. Illustrative examples were from RS and MA-A whilst MJ and SA-A provided valuable discussions. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Nottingham Geospatial Institute, University of Nottingham, Nottingham, UK. [2]Plant and Crop Sciences, University of Nottingham, Nottingham, UK. [3]Plant Sciences, Southern Cross University, Lismore, Australia. [4]Faculty of Agricultural Sciences, Sabaragamuwa University, Belihuloya, Sri Lanka. [5]Crop For the Future (CFF), Semenyih, Malaysia. [6]Global Open Data for Agriculture and Nutrition (GODAN) http://www.godan.info.

References

1. Al-Azri M, Leibovici DG, Karunaratne A, Meek S, Ray RV. Simulating eyespot disease development and yield loss using APSIM for UK wheat. In: Agriculture and climate change, adapting crops to increased uncertainty, Amsterdam, The Netherlands, 15–17 February 2015; 2015.
2. Angulo C, Rötter R, Trnka M, Pirttioja N, et al. Characteristic 'fingerprints' of crop model responses to weather input data at different spatial resolutions. Eur J Agron. 2013;49:104–14.
3. Azam-Ali SN, Aguilar-Manjarrez J, Bannayan M. A global mapping system for bambara groundnut. In: FAO, agricultural information and knowledge management papers, vol. 1; 2001. p. 55.
4. Azam-Ali SN. Fitting underutilised crops within research-poor environments: lessons and approaches. S Afr J Plant Soil. 2010;27(4):293–8.
5. Basso B, Ritchie JT, Pierce FJ, Braga RP, et al. Spatial validation of crop models for precision agriculture. Agric Syst. 2001;68(2):97–112. doi:10.1016/S0308-521X(00)00063-9.
6. Batchelor WD, Basso B, Paz JO. Examples of strategies to analyze spatial and temporal yield variability using crop models. Eur J Agron. 2002;18(1–2):141–58.
7. Bigagli L, Santoro M, Mazzetti P, Nativi S. Architecture of a process broker for interoperable geospatial modeling on the web. ISPRS Int J Geo Inf. 2015;4:647–60. doi:10.3390/ijgi4020647.
8. Birch ANE, Begg GS, Squire GR. How agro-ecological research helps to address food security issues under new IPM and pesticide reduction policies for global crop production systems. J Exp Bot. 2011;62(10):3251–61.
9. Butenuth M, Gösseln GV, Tiedge M, Heipke C, Lipeck U, Sester M. Integration of heterogeneous geospatial data in a federated database. ISPRS J Photogrammetry Remote Sensing. 2007;62(5):328–46.
10. Castronova AM, Goodall JL, Elag MM. Models as web services using the open geospatial consortium (OGC) web processing service (WPS) standard. Environ Model Softw. 2013;41:72–83.
11. Cruz SAB, Monteiro AMV, Santos R. Automated geospatial web services composition based on geodata quality requirements. Comput Geosci. 2012;47:60–74.
12. Dorigo WA, Zurita-Milla R, de Wit AJW, Brazile J, et al. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modelling. Int J Appl Earth Obs Geoinf. 2007;9(2):165–93.
13. Estes LD, Bradley BA, Beukes H, Hole DG, et al. Comparing mechanistic and empirical model projections of crop suitability and productivity: implications for ecological forecasting. Glob Ecol Biogeogr. 2013;22(8):1007–18.

Leibovici *et al. Open Geospatial Data, Software and Standards* (2017) 2:20

Page 11 of 11

14. Ewert F, Rötter RP, Bindi M, Webber H, Trnka M, Kersebaum KC, Olesen JE, van Ittersum MK, Janssen S, Rivington M, Semenov MA, Wallach D, Porter JR, Stewart D, Verhagen J, Gaiser T, Palosuo T, Tao F, Nendel C, Roggero PP, Bartošová L, Asseng S. Crop modelling for integrated assessment of risk to food production from climate change. Environ Model Softw. 2015;72: 287–303.
15. Frank A. Analysis of dependence of decision quality on data quality. J Geogr Syst. 2008;10(1):71–88.
16. Gaber N, et al. (2008) Integrated modeling for integrated environmental decision making, white paper EPA100/R-08/010, EPA (united states environmental protection agency).
17. Guillot G, Leblois R, Coulon A, Frantz AC. Statistical methods in spatial genetics. Mol Ecol. 2009;18(23):4734–56.
18. Hintum van TJL, Knupffer H (2010) Current taxonomic composition of European genebank material documented in EURISCO. In: Plant genetic resources: characterization and utilization 8 (2)S: 182–188.
19. Holzworth DP, Huth NI, deVoil PG, Zurcher EJ, et al. APSIM – evolution towards a new generation of agricultural systems simulation. Environ Model Softw. 2014;62:327–50.
20. Jackson MJ, Hobona G, Bernard L, Brauner J, Higgins C. The role and development of a persistent interoperability test bed for geosciences research. In: Sinha AK, Arctur D, Jackson I, Gundersen L, editors. Geological Society of America, societal challenges and GeoInformatics special paper, vol. 482; 2011. p. 69–75.
21. Kuhn W. Core concepts of spatial information for transdisciplinary research. Int J Geogr Inf Sci. 2012;26(12):2267–76.
22. Larsen TJ, Niederman F, Limayem M, Chan J. The role of modelling in achieving information systems success: UML to the rescue? Inf Syst J. 2009; 19(1):83–117.
23. Launay M, Guerif M. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. Agric Ecosyst Environ. 2005;111(1–4):321–39.
24. Leblanc MJ, Favreau G, Massuel S, Tweed SO, et al. Land clearance and hydrological change in the Sahel: SW Niger. Glob Planet Chang. 2008; 61(3–4):135–50.
25. Leibovici DG, Hobona G, Stock K, Jackson M. Qualifying geospatial workflow models for adaptive controlled validity and accuracy. In: IEEE 17th international conference on GeoInformatics 2009. Fairfax; 2009. p. 1–5.
26. Leibovici DG, Jackson MJ. Multi-scale integration for Spatio-temporal Ecoregioning delineation. Int J Image Data Fusion. 2011;2(2):105–19.
27. Leibovici DG, Pourabdollah A, Jackson MJ. Meta-propagation of uncertainties for scientific workflow Management in Interoperable Spatial Data Infrastructures. In: EGU 2011, European Geosciences Union. Vienna: General Assembly; 2011.
28. Leibovici DG, Pourabdollah A, Jackson MJ. Which spatial quality can be meta-propagated? J Spat Sci. 2013;58(1):3–14.
29. Leymann F. BPEL vs. BPMN 2.0: should you care? Lecture Notes Business Info Proc. 2011;67:8–13.
30. Liu J, Williams JR, Zehnder AJB, Yang H. GEPIC – modelling wheat yield and crop water productivity with high resolution on a global scale. Agric Syst. 2007;94(2):478–93.
31. Liu J, Folberth C, Yang H, Röckström J, et al. A global and spatially explicit assessment of climate change impacts on crop production and consumptive water use. PLoS One. 2013;8(2):e57750.
32. Loireau M, Leibovici DG, Desconnets J-C, D'Herbès J-M. LEIS, a tool for diagnosis and prevision of Antropogene pressure on natural vegetation. In: RGLDD (remote sensing and Geoinformation processing in the assessment and monitoring of land degradation and desertification). Trier; 2005.
33. Manel S, Holderegger R. Ten years of landscape genetics. Trends Ecol Evol. 2013;28(10):614–21.
34. Mayes S, Massawe FJ, Alderson PG, Roberts JA, Azam-Ali SN, Hermann M. The potential for underutilised crops to improve security of food production. J Exp Bot. 2012;63(3):1075–9.
35. Meek S, Jackson M, Leibovici DG. A generic BPMN client to enable seamless workflow composition of OGC web services. Comput Geosci. 2016;87:76–83.
36. Mirschel W, Wieland R, Wenkel KO, Nendel C, et al. YIELDSTAT – a spatial yield model for agricultural crops. Eur J Agron. 2014;52(Part A):33–46.
37. Molosiwa OO (2012) Genetic diversity and population structure analysis of Bambara groundnut [*Vigna subterranea* (L.) Verdc.] landraces using morph-agronomic characters and SSR markers. PhD in crop sciences University of Nottingham.

38. Morin E. Interdisciplinarité et transdisciplinarité. Transversales Sci Cult. 1994;29:4–8.
39. Nativi S, Mazzetti P, Geller GN. Environmental model access and interoperability: the GEO model web initiative. Environ Model Softw. 2013;39:214–28.
40. Neumann K, Verburg PH, Stehfest E, Müller C. The yield gap of global grain production: a spatial analysis. Agric Syst. 2010;103(5):316–26.
41. Olukolu BA, Mayes S, Stadler F, Ng NQ, et al. Genetic diversity in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) as revealed by phenotypic descriptors and DArT marker analysis. Genet Resour Crop Evol. 2012;59(3): 347–58.
42. Priya S, Shibasaki R. National spatial crop yield simulation using GIS-based crop production model. Ecol Model. 2001;136(2–3):113–29.
43. Reynolds JF, Grainger A, Stafford-Smith DM, Bastin G, Garcia-Barrios L, Fernandez RJ, et al. Scientific concepts for an integrated analysis of desertification. Land Degrad Dev. 2011;22:166–83.
44. Recker J. BPMN research: what we know and what we Don't know. In: Mendling J, Weidlich M, editors. Business process model and notation. Lecture notes in business information processing. Berlin Heidelberg: Springer; 2012. p. 1–7.
45. Rounsevell MDA, Annetts JE, Audsley E, Mayr T, et al. Modelling the spatial distribution of agricultural land use at the regional scale. Agric Ecosyst Environ. 2003;95(2–3):465–79.
46. Santos R, Algar A, Field R, Mayes S. (2017) Integrating GIScience and Crop Science datasets: a study involving genetic, geographic and environmental data. PeerJ Preprints. 5:e2248v4. https://doi.org/10.7287/peerj.preprints.2248v4.
47. Sghaier M. Analyse de l'intégration entre les données socio économiques et écologiques: synthèse régionale. In: Requier-Desjardins M, Ben Khatra N, Nedjraoui D, Issoufou WS, Sghaier M, Briki M, editors. Surveillance environnementale et développement : acquis et perspectives - Méditerranée, Sahara et Sahel. Montpellier: CIHEAM (Centre International de Hautes Etudes Agronomiques Méditerranéennes), 2012. (Options Méditerranéennes, Série B : Etudes et Recherches, n. 68); 2012. p. 71–94.
48. Tan G, Shibasaki R. Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration. Ecol Model. 2003;168(3): 357–70.
49. Van Ittersum MK, Ewert F, Heckelei T, Wery J, et al. Integrated assessment of agricultural systems – a component-based framework for the European Union (SEAMLESS). Agric Syst. 2008;96(1–3):150–65.
50. Wainwright J, Mulligan M, editors. Environmental modelling: finding simplicity in complexity: Wiley; 2013.
51. Wilson P, Harpur N, Darling R. Explaining variation in farm and farm business performance in respect to farmer segmentation analysis: implications for land use policies. Land Use Policy. 2013;30(1):147–56.
52. Xiao Y, Mignolet C, Mari J-F, Benoît M. Modelling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: a case from France. Comput Electron Agric. 2014;102:51–63.