MULTIPLY (H2020-687320)

REPORTS OF KICK-OFF MEETING, PROJECT MEETINGS AND FINAL MEETING

DOCUMENT REF:	WP1_D1.3_D3
DELIVERABLE REF:	D1.3
VERSION:	v1
CREATION DATE:	2019-12-28
LAST MODIFIED:	2019-12-28

H2020 MULTIPLY Kick-off Meeting (final draft)
Leiden, 7 January 2016

Present: Alex Löw (LMU), Thomas Weiss (LMU), José Gomez-Dans (UCL), Philips Lewis (UCL), Carsten Brockmann (BC), Grit Kirches (BC), Lea Hallik (TO), Mait Lang (TO), Itziar Alonso Cañas (UAH), Jon Styles (Assimila), Nicola Pounder (Assimila), Bethan Perkins (Assimila), Daniel Kindred (ADAS), Philippe Peylin (LSCE), Natasha MacBean (LSCE), Amie Corbin (CML), Peter van Bodegom (CML, chair), Esther Philips (CML, minutes)

Introduction by Peter van Bodegom

1. **Who is who**
   All partners shortly introduce themselves.
   A picture will be taken later this afternoon to put on the project website.

2. **A. Grant Agreement**
   Time records should be kept on a monthly basis, also for staff that is exclusively appointed on the project. It is advised to directly register the hours per work package, since the distribution of time spent on the separate work packages should be reported to the EU.

   Reporting periods: M12, M30 and M48.
   Reports should be uploaded maximally 60 days after the set period.
   A reminder and templates will be provided (Esther).

   An overview of when, where and how to acknowledge the EU will be sent by email. (Esther)

   Once the consortium decided on a logo for MULTIPLY, also a PPT template will be distributed (Esther).

   Although open access is obliged, the exact rules will be checked after which the consortium will be informed (Esther).

   B. Consortium Agreement
   The Consortium Agreement has been signed by all partners and will be countersigned after which it will be distributed.

3. **Financial procedures and issues**
   A notice of receipt of the pre-payment will be sent to all partners (Esther).
   The threshold for an EU audit will be checked (Esther).
   It will be checked if there are any consequences if more or less time is spent per WP (Esther).

   The intention is to work with a Project Management Tool. Any experiences or tips (besides Teamwork.com and EMDESK) are welcome (all to Esther)

   Esther will check whether external reviewers will be involved in the project evaluations.

4. Management structure: Executive board & Advisory board

The **Executive Board (EB)** consists out of the Work Package leaders, completed by 1 representative per partner. The list as suggested is agreed upon:

1.	Peter van Bodegom	Leiden University
2.	Alexander Loew	Ludwig-Maximillians Universität München
3.	Philip Lewis	University College London
4.	Carsten Brockmann	Brockmann Consult GmbH
5.	Lea Hallik	Tartu Observatory
6.	Itziar Alonso Canas	Universidad de Alcala
7.	Jon Styles	Assimila
8.	Daniel Kindred	ADAS UK Ltd.
9.	Philippe Peylin	Université de Versailles Sant-Quentin-en-Yvelines

All members will identify a back-up person and send name and contact details to Esther.

The EB will have monthly tele-conferences (60 minutes) where progress in general is discussed for 45 minutes, leaving 15 minutes for a presentation by one of the partners on science (related to the project but not necessarily part of the project). The presentation will be distributed before the meeting.

A doodle will be sent to plan ahead (Esther).

Advisory Board (AB)

The suggested AB is based on the projects' supporting letters. It should consist of users and potential users. Besides the current list, representatives of the following organisations/consortia and/or the names mentioned may be added to the AB:

- ESA, Michael Berger
- MPI Jena.
- A BACHY representative
- Person from JRC working with biodiversity from remote sensing (more ecology or ecosystem background)
- Marc Patanini
- Steven Plummer
- Benjamin Poulter

The suggested members are not specifically users themselves, but they can distribute project results.

Peter will contact the proposed members to invite them as AB member. The consortium will be informed on their response (Peter/Esther).

Frequency of involvement: 1x per year for ½ day after the regular project meeting.
The AB will be invited for the 1st meeting in M6 in combination with the user requirements meeting scheduled then.

5. Planning of all project meetings: when & where

when	partner	country
June/July 2016	LMU	Germany
January 2017	BC	Germany
June/July 2017	UCL	United Kingdom
January 2018	UAH	Spain
June/July 2018	TO	Estonia
January 2019	LSCE	France
June/July 2019	ADAS/Assimila	United Kingdom
Final meeting (Dec 2019)	UL	NL/Brussels

6. Project websites & logo

Following the workplan 3 web environments with each its own audience are to be delivered:

- E-learning environment (LMU)
- Platform (BC)
- Project website (UL); providing general info: partners, aims, key results or links to publications, links to platform & e-learning environment.
  URL: Multiply-H2020.eu

A project logo is considered useful and will be designed (Peter).

7. How to maximise societal impacts?

The Platform and E-learning environment should at least address key users as identified. Besides the AB members, Planet Lab is a key user with whom to stay in contact (Philip Lewis).

Workshops

D9.1 is a key user requirement document, to be delivered in M12, while M2 promises a user requirement workshop. After some discussion it is ***decided to organise a user requirement workshop at LMU next to the project meeting in M6.***

The date should be fixed a.s.a.p. and the key-users should be invited (and was fixed at the end of the 2nd day). Besides, at least 1 or 2 companies should be invited. It may be good to

check interesting contacts through the Spanish Ministry, since they have a web dissemination service (José). Through JRC it may be possible to identify individual biodiversity users that need to be invited to the workshop.

The invitations will be send back-to-back with the AB invitation (Peter/Esther).

A next workshop is planned in M42. This will be discussed at a later stage.

The summer/winterschool to be organised in M45, maybe to be combined with ESA activities, e.g. add a day to the ESA data simulation workshop or organise a joint ESA/ MULTIPLY Summerschool, hosted by ESA. The planning by ESA (Pierre-Philippe) should be checked two years in advance to know when to act (Peter).

E-learning environment
José gives a presentation on 2 examples. Important issues to consider are:
- The E-learning environment should work on the cloud.
- How to truly involve users, also afterwards (continuous involvement)?

If you make the E-learning environment too abstract and generic, people lose interest.
But if you make it very concrete, you may only reach a very small group. Ideally, however, it should apply to any user who wants to use SENTINEL data. In that respect, we will have three categories of users. 1. The scientists who want to use the platform for their science (most of the current training is dedicated to them). 2. Representatives of the advisory board and similar agencies. They will bring the platform into their own environment and may act as intermediary to end users. The user requirement workshop should clarify their demands, for both platform and e-learning environment. 3. End users. They will probably not use the platform directly, but may have to be trained in how to interpret its products.

The first example shows an E-learning environment designed for any user that wants to use SENTINEL data. It teaches how it works in general and gives examples. Users can discuss and exchange with other users or developers, while the environment can also be used as communication and guidance platform. In this way, examples in the e-learning environment may grow organically.
The other example concerns a training environment for using the platform. Experiments are explained with sample data. This way you slowly learn about real applications on the Platform. The E-learning environment is meant for testing, training and development and may not be directly connected to the Platform. Real problems should be run on the Platform, with the processing chain should be linked to the Platform.

A lot of issues are foreseen and discussed. The discussion will continue on the 2nd day of the Kick of Meeting, where it is on the agenda.

Platform
The interface of the platform will need more extensive discussions. Also, the server needs attention: Is the platform developed for one specific environment (e.g. SEMS) or should users be able to adapt the code to connect to any server of their choice? Also users will have to be prepared that running the platform takes considerable time. Independent of these decisions, the calculations need to be speeded up.

Action overview

No.	What	Who	When (deadline)
A2016-001	Provide reporting template to all partners	Esther	M10, M28, M46
A2016-002	Send overview of when, where and how to acknowledge the EU	Esther	12-02-2016
A2016-003	Send PPT template to all partners	Esther	01-03-2016
A2016-004	Check exact rules on open access	Esther	01-03-2016
A2016-005	Send notice of receipt of pre-payment to all partners	Esther	08-01-2016
A2016-006	Check the threshold for EU audit	Esther	01-07-2016
A2016-007	Check if there are consequences if more or less time is spent per WP	Esther	01-07-2016
A2016-008	Identify back-up person for Executive Board and send name and contact details to Esther	All EB members	15-02-2016
A2016-009	Doodle to plan telecon time slots with EB	Esther	08-02-2016
A2016-010	Contact and invite proposed Advisory Board members	Peter	01-03-2016
A2016-011	Plan 2nd project meeting at LMU	Esther	01-03-2016
A2016-012	Get design for project logo	Peter	01-03-2016
A2016-013	Contact Planet Lab to get them involved as key user	Philip	01-04-2016
A2016-014	Plan a user requirement workshop next to 2nd Project Meeting (M6)	Esther	01-03-2016
A2016-015	Contact Spanish Ministry to check interesting contacts (E-learning Environment)	José	01-04-2016
A2016-016	Invite key-users to requirement workshop	Peter, Esther	01-04-2016
A2016-017	Check planning ESA for summer schools	Peter	01-07-2016
A2016-018	Apply for domain name Multiply-H2020.eu	Esther	08-02-2016
A2016-019	All presentations given, as well as the minutes of the meetings and an overview of all actions to be taken will be distributed.	Esther	
A2016-020	Contact ESA in order to receive extra Copernicus Data. Does anyone know a better way to attain Copernicus Data than through the Science Portal?	Peter	
A2016-021	Check if EU external reviewers will be involved in the project meetings.	Esther	

Decisions taken

No.	What
2016-001	Members of the Executive Board decided upon
2016-002	Organise a user requirement workshop at LMU next to project meeting in M6

H2020 Multiply Kick-off-Meeting (final draft)
Leiden, 8 January 2016

Present: Alex Löw (LMU), Thomas Weiss (LMU), José Gomez-Dans (UCL), Philips Lewis (UCL), Carsten Brockmann (BC), Grit Kirches (BC), Lea Hallik (TO), Mait Lang (TO), Itziar Alonso Cañas (UAH), Jon Styles (Assimila), Nicola Pounder (Assimila), Bethan Perkins (Assimila), Daniel Kindred (ADAS), Philippe Peylin (LSCE), Natasha MacBean (LSCE), Amie Corbin (CML), Peter van Bodegom (CML, chair), Esther Philips (CML, minutes)

1. Goals and responsibilities

The proposal is ambiguous in its promises. In the proposal it reads:
"The project will develop an efficient and fully traceable platform that uses state-of-the-art physical radiative transfer models, within advanced data assimilation (DA) concepts, to consistently acquire, interpret and produce a continuous stream of high spatial and temporal resolution estimates of land surface parameters, fully characterized."
Possible route: we commit to develop platform software and we demonstrate that it works and we will show how it can be used and implemented. Since the product that is derived by the platform is complicated, it will not be a tool that is simple in processing.

In the main technical objectives, two aspects of delivery are highlighted:

1. *To provide users with data products for their **region of interest**, thereby bringing local mapping to the human scale for operational service use.*
2. *To fully employ **high resolution data** by providing high resolution data products, **constrained by additional low resolution** information.*

This does not necessarily imply a service. However, we should also keep in mind that the call text indicated that "a significantly wider use of SENTINEL data should be achieved". What is the best way to reach this? How will we make sure people keep using the data? The best way forward seems to be to deliver software to be used later on and is easy to run and install.

One of the biggest challenges we face to do so, is to speed up calculations (now with emulators it takes 1.5 ms to run it for one location). Another potential way to speed up calculations is the development of a new solver. The current solver spends a lot of time to find the solution near the very end. This can be improved.

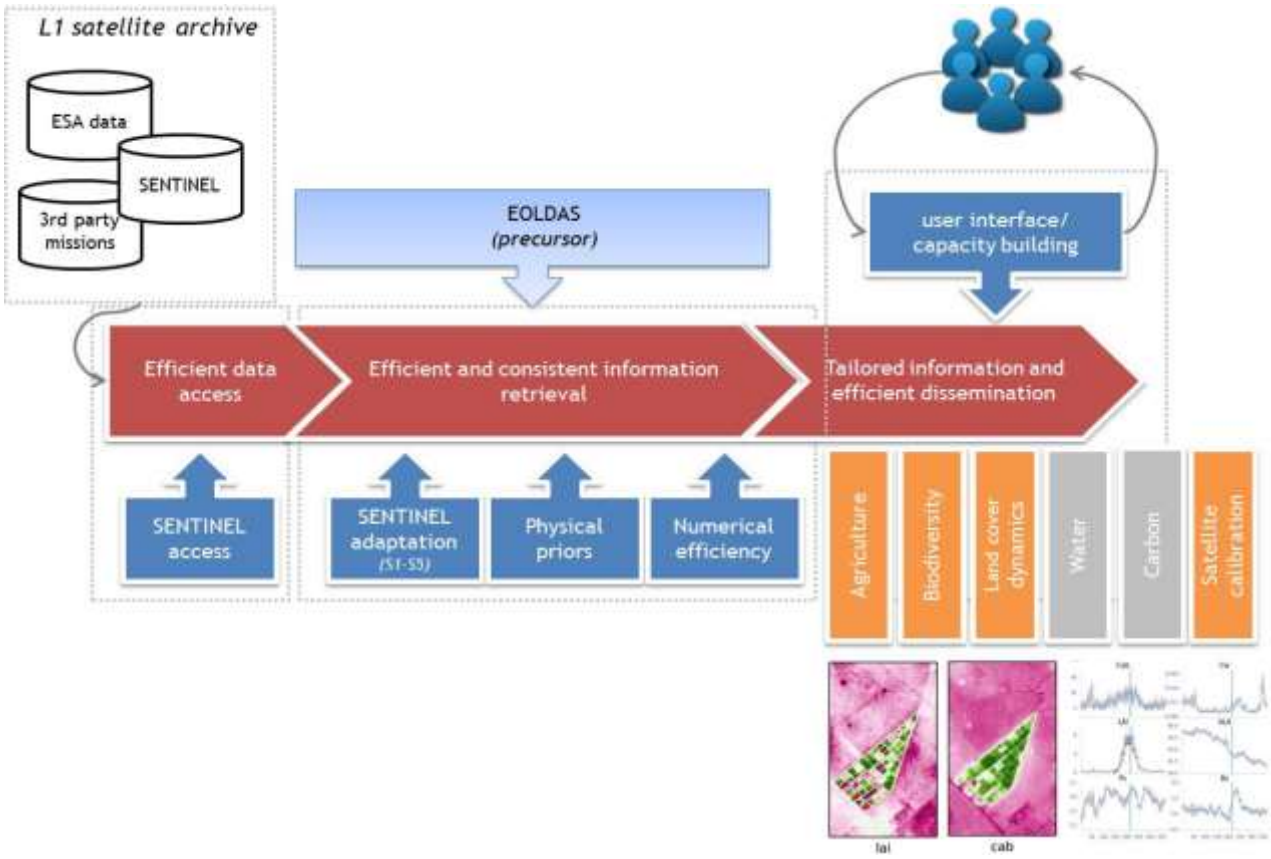
During the meeting with the external user group, an inventory of special interests, e.g. for a specific dataset, and requirements must be prepared.
Looking at the main policy relevant objectives:
The 5th (*To **facilitate** the **development** of **future Copernicus land surface products** and to specifically aid the **development of Essential Climate Variables**)* is tricky since we don't do a lot more than providing a platform
The 6th, 7th and 8th are covered in the demonstrators.

6. *To **stimulate** the **use** of EO-data products for climate, land use and vegetation modelling by providing consistent products*
7. *To **help evaluating policy measures and scientific developments** that aim at optimising crop yields within sustainable agricultural systems.*
8. *To **support** the **development** of **better satellite products***

The 9th is covered in the platform, e-learning environment and workshops.

9. *To **build** the **capacity of users** to retrieve, explore and work with EO-data through teaching, e-learning tutorials and user demonstrator projects.*

2. Current state-of-the-art

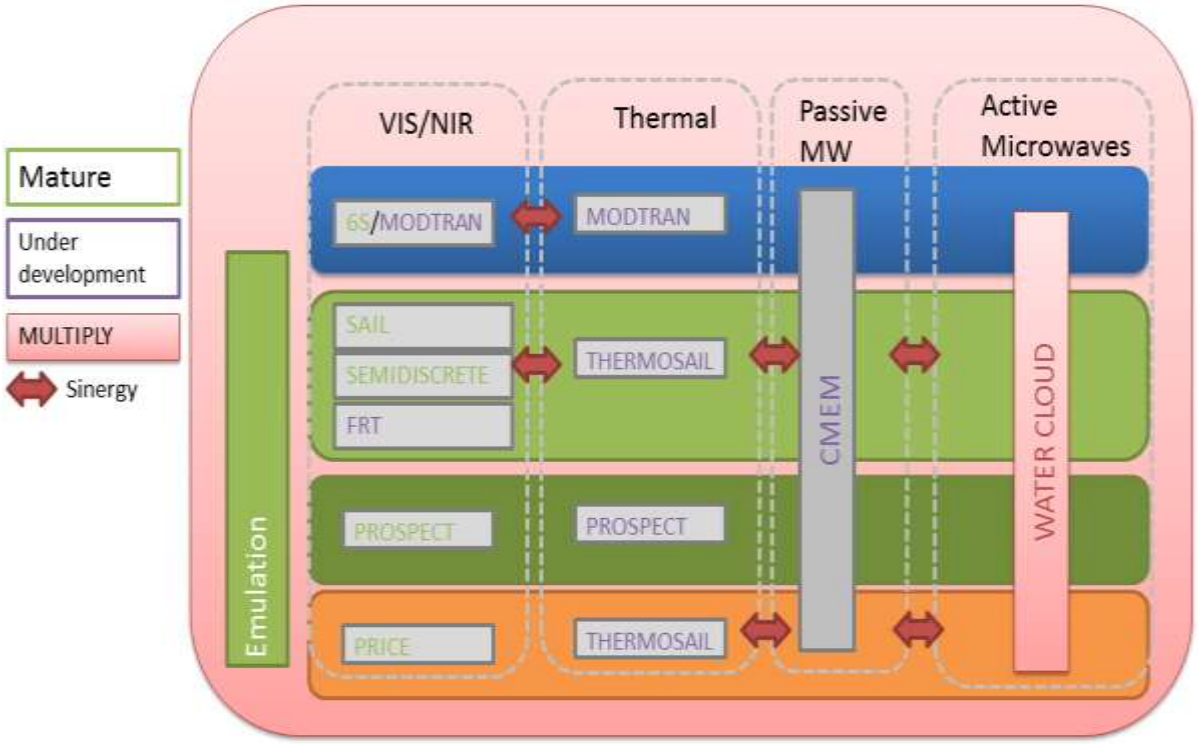


MULTIPLY schematic representation

User demonstrators are covered in the orange boxes. The left red box shows the internal platform, while the right red box shows the external platform. The user has to specify the problem to solve and sees the right red box only.

The figure down at the bottom right is to stress that the outcome is an interpretation of signals. It concerns spatial data processed over a specific field, followed by time series of different parameters. The novelty of MULTIPLY that it provides a full spectral picture of this reality (whereas albedo or MODIS LAI is not). Therewithit delivers the ability to un-mix what one doesn't want to interpret, leaving only those things that one is dealing with.

The deliverables we promised is a long list. If we identify a deliverable we consider to be not very useful, we should discuss this and also argue with the EC why we propose not to deliver it.



MULTIPLY predecessors

Different models have been proposed and implemented. In the proposal we have used CRMF. The forest and multilayer model is more or less ready to go. The thermal is only in the schedule for completeness and has not been promised in MULTIPY. A critical issue is, that all models describe things/scenes in a different way, for instance with respect to LAI. To ensure consistency among these scenes for different components is an important part of the project.

To speed up calculations, emulators have been made. You only need to make an emulator once and you may think of emulators as look-up tables. Hence, it is a two-step approach. The emulator needs to be built once and then can be used forever.

Based on the data assimilation, posterior uncertainties are provided (based on the difference between model and reality). The model uncertainty is not being accounted for. The user is supposed to be able to explain the choice of the model and to know the best model to use. However, if you think a particular model is better, you can emulate it and put it into the platform.
You can run multiple scenarios and find out which model works best, but doing this for model selection is tricky. The user is supposed to set up the problem.

The scope of WP4 (pre-processing) needs to be decided upon. The question is whether the input to the platform is top of atmosphere irradiance. For coarse resolutions, a product is in principal available. The question is how to treat high resolution data.

The current set-up assumes gradual changes. An improvement in the scheme has to do with the detection of disturbances. UCL has made progress on this, and these have been tested in

time, but not yet in space (while e.g. deforestation is about abrupt changes in both space and time). Snow, another abrupt change, is easily detected/ Detecting the moment of melting is more difficult as it implies a gradual change.
Coarse resolution data help as background reflectances to help interpreting high resolution data (and allows filtering things you don't want to interpret such as roads).

3. What did we promise?

From the Call:
Stimulate wider **research** use of Sentinel data. What we deliver should be useful (upstream).
Scope: tools and access methods need to be developed.
Expected impact: Copernicus Sentinel data used more.

From the DoW MULTIPLY:
Platform
Framework to exchange components as plug-ins
Develop **infrastructure**
We promise some **products**.
We mention **user interfaces**
Altogether, this implies that we need the semantics and level of the product right. This will have to be an important topic for the next meeting.

> If we want to deliver software it should be made configurable, easy accessible and it should have an interface that allows others to use it.

Platform:
3 scenarios:

1. Consortium runs the platform
2. Or pack it in virtual machine for scientists to use
3. Or ESA hosts a virtual machine (in which case the platform is tailored to specific hardware), but for this we should know if they want to host it.
   (Or Copernicus landholder could host virtual machine – and convince the EU to pay for the service).

After a discussion it is concluded that scenario 2 is the preferred one (the question is what the key users have in mind).

The Platform doesn't include hardware as deliverable but we need hardware to run it. We are obliged to get at least one instance of a virtual machine running.

4. Data Management Plan

A Data Management Plan should be delivered in Month 6.
Esther distributes a template used at UL.

5. Validation cases

Although linked to demonstrators, validation cases are different. Demonstration should be kept in mind though when defining validation cases, e.g. to include agricultural sites.

1. Performance of the retrieval using synthetic datasets from the QA4ECV programme

2. Validation "for a number of selected sites"
   a. Jarvselja (Estonia, Lea)
   b. ESA SENTINEL validation campaigns,
   c. ESA datasets, candidate site Barrax (= agricultural site)
   d. Collaboration with EnMAP campaigns, JECAM (also agricultural aspects) datasets

   Not only use existing datasets because they don't use SENTINEL data.

We deliver a preselection of metadata in M12. Sooner would be better in order to have some actual data. For that purpose, a document will be created (initiated by LMU) to which people can add potential additional sources.
For such metadata compilation, it is essential to define the state factor.
To have an impact on campaigns taking place this year, and to decide whether taking extra measurements makes sense, we need state factor at the end of February, beginning of March.
UCL will send a Jena document to the partners on this state factor.

6. User demonstrators

Data management on data used within the project should be decided upon (internal database); WP5 (meta data plan): we should agree who does what, who is responsible. The most obvious partner for a common repository of the information is UCL.

Types of data concerned:
- Satellite data: storages space, access, etc.
- In-situ data. Has different file format etc. Think about design of in-situ data such that is quality controlled and can be coupled to satellite data.

A Repository will be set up at UCL for all data.
The test case to start with will be the Estonian site.

3 user demonstrators are identified:
Modellers: LSCE
Agriculture: ADAS
New satellite products: e.g. enMAP

<u>LSCE:</u>
- Application in DGVM
- Use: global model, different vegetation types.
- Important for LSCE: land management for which simulations at higher resolution are required. MULTIPLY data can help on this.
- The target is to perform global simulations, but what and how is not exactly decided upon yet.
- MULTIPLY data can deliver the input.
- A new feature is the multi-layer canopy scheme (energy, water, carbon balance measured) and turbulence of canopy within the scheme.
- In future: simplified model.
- Important for LSCE = turbulence within canopy. Hopefully MULTIPLY can help.
- Critical info: sparse against dense vegetation.
- Six sites that LSCE already uses can be used. Try to include at least some of these sites within validation.

- LSCE also uses a data assimilation system and the 2-stream model.

Philippe will send the paper on implementation of the 2 stream model.
Philippe: share sites (also the crop sites)

The Model can be used for the Estonian site also.

Questions:
Q: Are there any specific crop sites that LSCE works on as well?
A: A developed version of the model (STYX) includes crop sites.

Q: What info would be good to share to prepare for the user requirement workshop in order to get the best possible outcome?
A: Write a list of realistic possible output (= state factor document)
Think about vertical profiles (multiple layer canopy models)

ADAS (Agriculture)

- ADAS' main satellite service used, are commercial systems (SOYL Sense)
- Agriculture uses of EO is very much used for policy/monitoring:
    - Crop type
    - Predicting yield
    - In field variable rate management
- ADAS would prefer a crop intelligence system
- The benefit comes from comparison; create curves robustly.
- Yield Enhancement Network (www.yen.adas.co.uk)
- It is crucial to understand and analyse yields.
- Link science with industry: Learn to speak the same language.
- Structure is used to estimate (bio physical) potential yield.
- The idea is that 'we' could do better (higher yield?).
- Farmers don't think about the real resources (e.g. water, soil), but they do think about BM (chemicals), etc.
- Raise in production based on natural effects?
- Identify subset of fields to start with

7. Wrap-up

Deliverables for the 1st year (2016)

jan-mar	apr-jun	jul-sept	oct-dec	what;who
D1.1				project website; UL
	D1.2			data management plan; UL
			D2.1	database optical obs operators; UCL
			D3.1	ATBD land surface priors; UL
			D4.1	ATBD pre-processing coarse; BC
			D5.2	meta-data catalogue time series; UL
D6.1				TN software dev. Env.; BC
		D6.3		system architecture doc; BC
			D7.1	prototype vegetation char. Database; UL
			D8.1	vision doc; BC
			D9.1	key user requirements doc; LMU

The next project meeting will be combined with a User Requirement Workshop. It will take place from 11 to 13 July 2016 in Munich, Germany (LMU).

Minutes Consortium Meeting d.d. 11 July 2016

Present: Alex Löw (LMU), Thomas Weiss (LMU), José Gomez-Dans (UCL), Carsten Brockmann (BC), Grit Kirches (BC), Tonio Fincke (BC), Lea Hallik (TO), Jon Styles (Assimila), Nicola Pounder (Assimila), Bethan Perkins (Assimila), Daniel Kindred (ADAS), Philppe Peylin (LSCE), Natasha MacBean (LSCE), Peter van Bodegom (CML, chair), Esther Philips (CML, minutes)
Absent: Philip Lewis (UCL), Tom Block (BC), Itziar Alsonso Cañas (UAH)

1. Software Development Plan

Carsten presents the software development plan, which is then discussed.

Year 1 (2016)

1. Develop the overarching architectural approach (high level) for the MULTIPLY Platform
2. Implement concept demonstrators
3. Re-use EOLDAS and other existing components
4. Postprocessing Vegetation types functional fingerprint; file interface with platform

1 release end of 2016 (MP v0.1, internal to team)

Year 2 (2017)

1. Develop data access components, specifically for S1, S2, S3
2. Migrate & upgrade EOLDAS Emulation framework and assimilation engine into MULTIPLY platform
3. GUI components & visualisation (initial)
4. Postprocessing vegetation types & biodiversity indicators: file interface with platform

2 releases mid 2017 (MP v0.5, internal to team) & end 2017 (MP v0.6, public beta 1 release)

Year 3 (2018)

1. Collecting feedback from team, decide on upgrades and improvements
2. Testing with Sentinel data
3. Gui components & visualisation (cont., low prioroty)
4. Postprocessing all components (vegetation types, biodiversity indicators, disturbance indicators)
5. Postprocessing framework integration (file interface & seamless integration)

2 releases mid 2018 (MP v0.9 public beta 2) & end 2018 (MP v1.0, public version 1)

Year 4 (2019)

1. Platform robustness (debugging, supporting demonstrations)
2. Gui components & visualisation (finalisation)

1 release (MP v2.0, final public)

2016:
The work done by Tom and Tonio in 2016 will be a learning curve of which knowledge can be used in the years to follow. The MULTIPLY Platform (MP) v0.1 release will contain components, including microwave components, which we know work and are meaningful to test (at coarse resolution and aiming at integration with high resolution data, such as Sentinel 2) and dummies for other components to be added later. The release will be demonstrated to the Consortium members only and the performance should be integrated in the architecture of the Platform.

2017:
This period will be used to build a first new version of the MP that should, in an operational sense, be close to the end result. This means Sentinel 1, 2 and 3 data streams can be used. For this it is necessary that data is organised in smart ways to enable easy retrieval. Developing these data matrices – as internal part of the platform- will be the most important challenge which will require close collaboration. An early stage approach is developed by UCL which should be improved at the end of 2017. Mid 2017 MP v0.5 shall be ready for internal demonstration

and testing, while MP v0.6 will be the first public beta release. This beta release will be available to 'friends and family'.

In 2018 the public beta MP v0.9 will be released, followed by MP v1.0 as first public version. In the autumn of 2019 the final public MP v2.0 will be ready.

It is important that all versions are tested by users, at a minimum the users within the Consortium.
A broader user group will be invited to our meeting at the end of 2017 when the first public beta MP version will be released.

Actions and conclusions resulting from the discussion:

1. Peter and José will be in touch with UAH on their role.
2. The required speed of processing data through the MP should be defined by the end of 2016.
3. The critical part the coming 6 months is to get the data structure/model right since this has a lot of implications on all other aspects within the project.
4. BC will take the lead in getting the input data right (coming ½ year)
5. Communication among the various WPs, also those now loosely connected, must be arranged and maintained

2. Data Management Plan (DMP)
There are no general comments to the Data Management Plan.

In the DMP 4 types of data are mentioned:

- Derived EO data
- Field validation data
- Raw EO data to be used within MP
- Field data from 3rd parties

We take our hands off the raw EO data of others, but do we indeed assume that to be the users' responsibility?
We are free to release converted raw data (no legal restrictions) but is that something we plan to do? Or will we only keep the converted raw data on an own server and show the user the derived data only?

Discussion results in the following decision:

- A short policy note will be drafted for the use of each date type. This should include a clarification which data we are allowed to provide or redistribute and which we can't.
- Data that we are allowed to use free of legal restrictions will be shown in a menu within the platform. In addition, the option will be given for a user to add his/her own data source.
- The DMP should explain the legal restrictions to the various data types.
- Options will be given in the user interface.
- For existing data to be used, the legal restrictions will be documented.
- Users of the MP must be registered users to the data sources they intend to use within the MP.

BC shows an example that would be a good option from a technical perspective. A lot of options are given so the output the user gets is answering his needs.

Next topic of discussion is how to connect derived EO data to the MP? During a Telcon it was decided that we want to archive the derived EO data so the output is saved and known. This requires conscientious version management. How to deal with version management is decided upon in Deliverable 6.1. (Technical note on the software development environment). It would be an extra service to users because it could provide them to do work that has already been done. It would be attractive to visualize this archive: a map on which you can see where someone has been processing which kind of data. However, we carefully need to consider what our responsibilities towards the EC are versus what we would like to offer. Even more so, since an archive (storage space, labour costs) is not budgeted for. The 'light' option could be that the only thing archived is the metadata on what was already processed and by who. The user could then contact this party. Users that run data should however make the decision if they want the metadata published or not.

If the Consortium decides that certain datasets are of common interest to be kept, a Digital Object Identifier (DOI) will be generated. This will likely relate to either a. validation sets or b. large datasets (e.g. entire Europe) that will lead to major interests by many users. Only those datasets will thus be archived. All other data products are considered intermediate products and stored only locally.

The Code Base Repository is based on Github and this will not change once the MP is launched. Tonio will create a MULTIPLY Github repository environment to which components can be migrated. Migration will take place in the first half of 2017. The Github repository will be open access.
Rules on licences of components brought within the MULTIPLY repository will be checked (Esther), since MULTIPLY should be public.
After the end of the project, the platform is released to different users and responsibility ends. In addition, each MULTIPLY version will be given a DOI (at a CEMS/Pangea or similar cloud) to ensure accessibility after the project ends.

The DMP indicates that the version system of the data will follow the system of the codes. Is that feasible?
We will follow the NetCDF+ and CF-1.7 metadata standards. Following the questionnaires, some users have different requests (e.g. GEOTIFF). The software should be able to run both. Option could be given in the MP.

Should MP be a service or a virtual machine?
At the end of this year we must make sure to decide. What info do we need in the coming months to be able to make this decision?
A table should be drafted in which possible constraints and limitations are listed (Esther). We need to consider all levels of complexity. The pros and cons for each approach needs to be included. Since it is unrealistic to think that we as a group can run a service in the future, we need to get feedback from strong players on hosting the MP after the project ends (e.g. JRC, ESA, EEA).

A structure for the metadata of the field data needs to be developed. It would be good if these were aligned and checked –also by the modellers- to make sure that it contains all information

that may be needed within the platform. If not, then field protocols will have to be adapted to make sure all essential variables are measured. A next version of the DMP should contain this metadata structure. The field protocols in use will be distributed (Alex, José, Daniel, Lea). Also the Jena protocol will be included by Jose to align the various field protocols. Carsten indicates the existence of a document directed at how to structure metadata (attached: Fernandes et al.). At the end of the project, the field data will be published with a DOI in an international data repository.
Esther will check the options for a Virtual Research Environment (VRE) at Leiden University. Here a template to collect metadata could be stored and commonly used, as well as all other internal consortium documents, including field data for internal use. The VRE will be used instead of a wiki or intranet.
Other actions following the 11th:

- The website needs to be updated to better reflect what MULTIPLY distinguishes from other platforms
- We need to start thinking about valorisation and follow-up projects by the end of year two.
- We need to be aware of the publication of a ECWMF call in the near future.

Minutes Consortium Meeting d.d. 12 July 2016

Present: Alex Löw (LMU), Thomas Weiss (LMU), José Gomez-Dans (UCL), Grit Kirches (BC), Tonio Fincke (BC), Lea Hallik (TO), Jon Styles (Assimila), Nicola Pounder (Assimila), Bethan Perkins (Assimila), Daniel Kindred (ADAS), Philippe Peylin (LSCE), Natasha MacBean (LSCE), Peter van Bodegom (CML, chair), Esther Philips (CML, minutes)

Absent: Philip Lewis (UCL), Carsten Brockmann (BC), Tom Block (BC), Itziar Alsonso Cañas (UAH)

1. Update actions & deliverables

- D1.2 DMP will be updated based on the outcome of this Consortium Meeting
- D6.2 should be delivered, even if only draft (BC a.s.a.p.)
- D6.6 system architecture requirement document UCL (due 30 September 2016)

2. Advisory Board (AB)

Current members:

Berger	Michael	European Space Agency
Guanter	Luis	Helmholtz Centre Potsdam
Poulter	Benjamin	NASA GSFC
Villa Alcázar	Guillermo	Ministerio de Fomento
Justice	Christopher	GEOGLAM (University of Maryland)
Lemoine	Guido	EU JRC
Bastin	Lucy	EU JRC

Additional members to invite

Contact for EEA*	Esther: Remind Jon to talk to Hans Dufourmont on 06-09-2016
PlanetLab	Esther: send invitation

Possible other interesting organisations

Carbon Map (Iain Woodhouse)	José: Send questionnaire and see how they respond
CATIE	Peter (given interest for REDD+ applications)

The best way to better involve the AB is to invite them to all end-of-the-year project meetings where the Consortium present the results up to date. This can also be in the form of a web-conference. Esther will check the technical options for this.

They will be invited to take part in the mid-term project meeting where the MP v0.6 beta public version is released (ideally to be hold either at JRC or UCL given e-infrastructure) as well as to the Summer school (4th year). Since during this meetings the MP will be demonstrated and the users will have the possibility to test, they should attend these meetings in person.

The AB members are the core of the foreseen users and will be expected to relate MULTIPLY to the rest of the user community. The members will receive a letter of what is expected from them (Esther).

What is also discussed is the need for up-to-date, informative and attractive information on the MULTIPLY website (for which links can be sent to the AB), including links and background information on predecessors and distinguishing added value for the global community (global datasets as developed in QA4ECV) and a tool for analyzing local high resolution data. This will be taken up by CML (Joris, Esther). In addition, info graphics (put on you tube) can be considered.

The best way to engage users is however to get a scientific paper out with the concepts and a proof of principle a.s.a.p.

3. **Questionnaires**
   So far we received approx. 12 responses and we expect to end up with around 20 completed questionnaires. If consortium members know of additional people whom can be contacted, they should do so directly with Esther in cc. We must make sure that the people who took the effort to respond, are also kept informed and involved.

4. **Field campaigns**
   The decision is made that a meta-data database will be created, containing required information on the various sites of previous and ongoing campaigns. This database will then be used to decide upon the sites to use for field campaigns. At this point it is at least clear that there is a strong preference for time series (i.e. 4-6 moments during the growing season) with multiple vegetation characteristics measured (and thus implying fewer sites to measure at). At the end of August, José will send a template.

   Efforts for next year should be coordinated and consortium members should be informed and aware of each other's field campaigns. An inventory of ongoing and starting campaigns should therefore be provided, including contact details. At the next project meeting it can then be decided who goes where, measuring what.

   Traceability chains will be implemented in the documentation of the platform.

5. **Next meeting**
   The next project meeting will be hosted by Brockmann:
   11 January 2017 from 13.00 – 17.30
   12 January 2017 from 9.00 – 13.00

Minutes Key User Requirement Workshop (part I), 12 July 2016

Present: Alex Löw (LMU), Thomas Weiss (LMU), José Gomez-Dans (UCL), Grit Kirches (BC), Tonio Fincke (BC), Lea Hallik (TO), Jon Styles (Assimila), Nicola Pounder (Assimila), Bethan Perkins (Assimila), Daniel Kindred (ADAS), Philippe Peylin (LSCE), Natasha MacBean (LSCE), Peter van Bodegom (CML, chair), Esther Philips (CML, minutes), Guido Lemoine (JRC), Lucy Bastin (JRC), Michael Berger (ESA)
Absent: Philip Lewis (UCL), Carsten Brockmann (BC), Tom Block (BC), Itziar Alsonso Cañas (UAH)

During the Key User Requirement workshop we will discuss the MULTIPLY Platform from the users perspective: what should the Platform ideally deliver in the sense of performance and variables?
After a short introduction on the MULTIPLY project by Peter van Bodegom, a presentation was given by José Gomez-Dans on what possibilities are already out there.

Presentation José Gomez-Dans:
There is the need to monitor the land surface from different perspectives:

1. Agriculture needs near real time tracking at high resolution
2. Carbon/Water/Energy (C/W/E) needs long time series and global datasets

The common requirements are:

- Interpret EO data in terms of land surface state
- Combine different sensors
- Consistent description of the land surface stage
- Uncertainty quantification

The measurements need to be interpreted through physical (RT) models that need to be robust and globally applicable. Data is acquired with different sensors to derive, e.g. heterogeneous spectral, angular, spatial, temporal properties. The RT models need to 'normalise' for these differences. However, data gaps, low sensitivity to some parameters, etc. is inevitable. MULTIPLY intends to cooperate with, for instance, EnMap. If we look at the current situation, we know that Sentinel data take very long to download and this will have to improve for real near time applications.

The full integration of Synthetic Aperture Radar (SAR) –or microwave data in general such as acquired by Sentinel I- is new to MULTIPLY and novel as no such system yet exists. Unfortunately though, microwave measures provide only a very indirect measurement of the specifics that are of interest to the researcher and the sensitivity to other occurrences in the surrounding area. There is a need for an adequate model to interpret the data. A key point is that at this moment the optical and the microwave range are completely decoupled where these will have to be integrated and combined in 1 platform. This implies that the internalities –such as LAI and measures of water contents that are in common to both- will have also be combined and the result would be a complete and consistent picture across. Of course, it is not as simple as putting things together.

Guido: SAR can be used to decrease uncertainty. When the focus is on agricultural data, the first thing to know is the crop type. After this is known, one can look at other aspects. It is of no use to develop an RT model without knowing the crop type.

José: different crop types would result in changes in the Leaf Angle distribution. Although optical images are of interest where vegetation is concerned, lots of changes go slow and are predictable. That
is why models can be used, although different crops ask for different priors. The ambition is to integrate all things and get a consistent result.

The challenges that MULTIPLY aims to tackle are:

- Speed up calculations to allow for near real time retrieval (and getting better phenology, post processing products, integration of microwave domain)
- Integrating codes into a Platform that allows users to retrieve data, access data archives and visualise
- Develop an e-learning environment with the aim to train as many users as possible.

An extra challenge might be to use MULTIPLY as a benchmarking tool for scientists. This may require that operational products are also integrated in the platform (or at least can be connected easily to platform outputs). That decision has not been taken yet.

The Copernicus operational service offers products and indices that can be used for (validation) benchmarking purposes. This is not the way MULTIPLY is foreseen; MULTIPLY is mainly foreseen as an analysis platform for users, and does not have as prime aim the provision of data products. Instead, we would like users to be able to plug their own algorithms as an alternative RT into the platform. In that case it will be possible to replace the RT scheme after running the emulation. This means an emulation framework should be provided within the platform for people to use their own RT. Some data products may be provided though, particularly for coarse resolution products.

Presentation Guido Lemoine (JRC, Ispra Italy)
MULTIPLY would be most relevant for the crop production monitoring part, but there is some potential in CAP control as well (e.g. grassland productivity at the longer term).

Concerning the commonly used EO data, data important to JRC is *Integrating crop area as a component in production estimates.* The variation within these crops is a next step of knowledge. There is some overlap with precision agriculture practices. This is an important field to gather information. Concerning the *Lower practical size limit around 1 ha parcel scale (with 10 m resolution,* anything below 1 ha is challenging. The *CAP controls (primarily VHR [< 1m], sample basis which is currently limited to area measurement and crop recognition)* is a closed environment. Current seasons are kept under the hat until the end of the season. It is interesting to know what S3 will deliver in addition of LAI, faPAR, impact of accumulation of biomass, stress indicators (i.e. the difference between actual and optimal LAI) and thermal data for evapotranspiration estimates

The most critical challenge seems to be having yet another platform. The European Commission is pushing for a solution to be set up and to bring the data to the processing and have intelligent solutions, ready to use by many users. One must be careful to come again with yet another platform. Technology is only interesting if the farmers gain something. JRC is interested in following the development of land properties over time (time series).

Presentation Daniel Kindred (ADAS)
The Industry does consider monitoring resources available to them. How much light is available? How much water is used?, etc.: that is the crop intelligent system ADAS wants to get to.
There is a great opportunity to better use time series of crop development
To allow the use of EO-data by farmers, near real time retrieval is essential, e.g. farmers want time critical information that cannot always be given. Such information is not only important for farmers, but also for technicians (e.g. the people hiring machinery to harvest, etc.), which are commonly forgotten.

There is an ongoing development in crop data to help farmers from a business side. It should therefore be considered what MULTIPLY wants to do. A general model is not enough since there are already specific crop models developed by the industry.

It would be great if other datasets were integrated within the MULTIPLY Platform.

Michael Berger – ESA
The European Space Agency is not a user as such, but a data provider.

If you develop a platform you will need to have well-defined user requirements at hand for different users. What the platform should provide is therefore not so difficult. A lot of platforms are developed in parallel. What is important is that the interfaces of all platforms should be able to 'collaborate'. This means a strategy is needed to fit together all activities that are going on. Multiply can play an important role here. In addition, multiply might for instance try to become the data assimilation part of the food security platform of ESA.

At the moment, a paradigm shift in EO is ongoing to make much better use of the EO-data integrated within a strong e-infrastructure. The ESA internal strategy is directed towards EO Innovation Europe, aiming for modern IT technologies on data assimilation and thematic platforms. There is no clear indication where the EC wants to go. The EC is not yet reacting to what is happening in the community or the market, but may release a tender on the Integrated Ground Segment (IGS) in autumn. This discussion should be taken up and developments followed.

MULTIPLY could provide a set of parameters as products to e.g. a decision system. A description of interfaces of what those platforms require is quite important which means a discussion on the interface for the MULTIPLY Platform is needed.

LSCE Philippe Peylin
The climate modelling community is merging more and more and there is an internal shift in the community towards taking notion of the need for consistency. If the community had a set of products that is consistent, this would be enough for people to take it up quite rapidly.

Another step, but a major challenge, is to get an emulated version of ORCHIDEE as a (RT) model into MULTIPLY. The complications arise from the fact that the RT-model of ORCHIDEE is strongly coupled to soil hydrology, phenology timing etc. and the various feedbacks detailed in the model.

Lucy Bastin (JRC-Digital Observatory for Protected Areas)
The relevance of the MULTIPLY Platform to this department of JRC would be the sharing of data ad information.
The main challenges are:

- Visualisation in itself – access to an Open Geospatial Consortium Web Map Service (OGC WMS is all we need. Visualization of uncertainty – needs more thought; Propagating and communicating uncertainty is important.
- Avoiding circularity: not all inputs to the products we use are described thoroughly and may be based on the same subset of products.
- Identifying salient changes early: nice to have early warning system but there is a limit to what can be done. Yearly probably e.g. to detect small logging roads/encroaching plantations
- Identifying ready for use global datasets that are consistent enough across time and space
- PAs vary hugely in size and nature.
- Drawing conclusions about the relationship between EO variables and our response variables is tricky
- PA boundaries and species ranges change over time, but often this reflects better data – no real change:
  - PA smallest: individual tree
  - PA largest: thousands of kilometres
  - PA is also political decision.

Is there an interest for using PA as case in MULTIPLY. Possibly yes. An example would be interesting. Such PA should have a good set of ground data and EO data and may be evaluated for detecting anomalies

Questionnaire: ***Q1 - products to obtain***
A list of other variables was given (next to basic ones) of which one is temperature. Thermal application has been left out because of the complications foreseen, especially in integration with the other components. If a user could plug in his own RT model, this may solve the problem. There are models available where it could be done and it would be a pity if the thermal part was missing. It is a valuable tool for simulation experiments too.

Most other variables mentioned are provided (e.g. phenology, soil moisture). What we don't provide is biomass related information and crop type. A discussion on what possibilities and requirements exist takes place of which the main issues are:

We need to know if we have a tree or grassland cover . Why not use ESA land cover?
Land cover could be used for the prior as well as to decide which RT to use. Should it be an independent output or an additional assumption in the platform? If we impose a land cover CCI product then we assume that land cover is true and you could make an assumption. However, if the land cover is wrong we blow up uncertainties in the product. On the other hand, if we calculate it internally as a product, it is questionable whether user will apply this product, given the availability of the land cover CCI product.

Different users have different needs. Ideally you would be able to pick up the anomalies,

Within the priors, are we defining priors per land cover type, or as narrowly as possible but not specifically? Or something different? We will try to be as precise but generic at the same time. Which level we will end up is not known yet.
Then: use the RT model (e.g. short vegetation) or plug in an own RT model if the user has a better one. There should be an automated system that makes a choice what RT model should be used.

We need flexibility. A technical solution would be: ability to read in a map (could be Copernicus). Then you have a table for selection of the RT model and priors that the user can find and change if needed. This way it is also clearly documented.

Uncertainties is one of our selling points and this is considered important.

The time required for near real time retrieval is completely different among the users. So: what should our target be? If the choice is made to provide a virtual machine, the operation speed depends on local services and servers. If the choice is made to provide a service, all data need to be mirrored in the MULTIPLY Platform. The speed of the processes is related to the choice that we make here.

Michael: You should review which user community you are aiming at. ESA standard is 3 hours after sensing but this is very costly.
For a small space mission it is now obliged that all earth explorer missions have real time. This doesn't count for the global field.

The issue should be thought about more carefully and the communities should be consulted: "What processing scheme do you want to consider?" Ask them what is the time lag now and see what magnitudes is given for improvement. Sequential retrieval is faster but results in more uncertainties.
Some back of envelope calculations are needed. We know the swath width of the Sentinels and can thus calculate the number of pixels released per units of time, giving a target for analysis. We can also identify how far off we are currently do determine how much faster is needed.

Minutes Key User Requirement Workshop (part II), 13 July 2016

Present: Alex Löw (LMU), Thomas Weiss (LMU), José Gomez-Dans (UCL), Carsten Brockmann (BC), Grit Kirches (BC), Tonio Fincke (BC), Lea Hallik (TO), Jon Styles (Assimila), Nicola Pounder (Assimila), Bethan Perkins (Assimila), Daniel Kindred (ADAS), Philippe Peylin (LSCE), Natasha MacBean (LSCE), Peter van Bodegom (CML, chair), Esther Philips (CML, minutes), Guido Lemoine (JRC), Lucy Bastin (JRC), Michael Berger (ESA)
Absent: Philip Lewis (UCL), Tom Block (BC), Itziar Alsonso Cañas (UAH)

1. Requirements for visualisation and data retrieval

The given set of visualisation requirements is considered to be the minimum required. A web-based visualisation needs discussion since the outcome of the questionnaires so far don't give a clear picture.

Lucy: Web based visualisation can be done with already available tools elsewhere if the data received is in the right format. It would be better to put the effort into combining already available datasets.

Archiving data would be really useful in order to retrieve the data in the future. What would be nice is the ability to click on a typical site (or preferably any pixel) on a map of which you then get the related data and other information. However, such archiving only works if the MULTIPLY Platform is offered as a service. If we would include other virtual machines, we should draft a catalogue and assure that the virtual machines can be connected and/or we would need to have a global run available as archive. This suggest the need for a central service. We could consider this in the future.

The users (ADAS & LSCE) are asked if they would prefer to download the data on their own server to continue visualisation and analysis (e.g. in ARCGIS or R environments) or to view and analyse the data directly in the MULTIPLY Platform, e.g. maps or time series graphs to analyse whether the output makes sense. Either way an overview of the downloadable formats is important

The focus of the ongoing discussion is too much on products. The objective of MULTIPLY is to produce a flexible tool and show something reasonable comes out of it. So, in terms of visualisation it is important for users to process and operate data on their own conditions.

Michael: on the issue of visualisation you have to consult the user base. What are their requirements?
Guido: Why worry? Better worry about how the models and validation fit in the particular model (tool?). If visualisation is needed, a small script could be used. It should all be patchable. If one can run a radiative transfer model in parallel relatively quickly (1 minute), that would be interesting. Output formats are not that important. What is not ready available can be made. The main frustration of a lot of users is that there are 50 different formats in GIS and in the end everybody wants GEOTIFF.
A totally new paradigm is needed. Platforms contain algorithms mainly to talk to the own organization environment. Therefore an infrastructure must be built to allow people to integrate data. MULTIPLY can deliver such infrastructure.

Alex: The point Guido makes is that of an expert user case view: at the back end the expert user can visualize by using a script. But not all users may want to do this or have the experience to do this. So, although not exclusive, we still need a web interface for less experienced users. A key point is the need to stick to standards for input and output. A critical condition is that APIs for the expert user should be updated in parallel to the web environment. You need an API that is Open Geospatial Consortium (OGC) like.

We need to communicate clearly that MULTIPLY provides an engine/ toolbox that will do the processing for you and will provide integration of different sensors to provide a state vector.

Uncertainties are not mentioned at all. MULTIPLY should make the difference here. Uncertainties are calculated and should be visualised at some scale.

Conclusion
The system needs to have 2 functionalities:

1. API for Expert users
2. Cloud web service for less expert users. Within this web service, we will have a stronger focus on the inputs; selection of ROI, selection of the priors and RTMs and output analysis seems less needed.

Given the availability of many other ‘’platforms’’ we may have to reconsider the name platform.

2. Sabine Chabrillat (user EnMAP and other)
The difference between EnMap and Sentinel is the number of bands. It is important how we caninterface with EnMAP. Brockmann is familiar with EnMAP although doesn’t have the latest information on algorithms in the EnMAP Toolbox. Details on this need to be checked with Sabine (Carsten). The tools seem to be complimentary but coming from a different philosophy.

We should continue discussions to see how we can connect, e.g. cross validation as MULTIPLY has the means to cross validate with other satellite products or bring in hyperspectral data and see what it adds to Sentinel data. They have direct access to hyperspectral data. That seems the most interesting way to exchange ideas.

3. Infrastructure and processing
Various possibilities.

	1 Service by team	2 Packed VM	3 EU Service	4. Regional service
Maximize use of Sentinel data (Call objective)	2	5	8	8
Comfort for team	4	9	1	4
Feasibility	6	9	1	4

Most users are able to run the Platform as a virtual machine, but they prefer a centralised service. This means a centralised sustainable solution after the project is desired. We start with Service by team/Packed VM and hopefully move to EU Service/regional service at a later stage.

Michael: What are the sustainability thoughts after MULTIPLY ends? Who will host the Platform and how? Financing will be needed for sustainability, maybe possibilities occur via a Call for Proposals (ESFRI).
Peter: If we can arrange hosting by another party, sustainability chances are bigger. Ideally we identify a party that wants to host the Platform as a service. It should by default be open access and open development. But also during the project the Platform must be hosted.
Carsten: Our purpose is to make tools available. We have to be open and find the best way for users to work with it. This means it can be running on a MULTIPLY Platform as a service or as a virtual machine. It might become part of the Copernicus services but this needs to be explored. The question is: whom to approach? Suggestions are welcome.

The main goal of the EC is that scientists should get tools to make better use of Sentinel data and the outreach should be broad. We therefore need to understand the working environment of the scientists.

Guido: A key point from the Sentinel 2 connection is that all data should be available corrected for the atmosphere. A toolbox approach is interesting because users can batch that. If that is a model, it's already very operational. The added value of MULTIPLY is having these kind of algorithms, because it goes clearly beyond Google Earth and so.

4. Extent to which data retrieval needs are fulfilled?
The tedious aspects come back again to the interface.
If there are suggestions for other communities to approach with the questionnaire, please provide names and addresses (e.g. microwave community) or contact directly (with cc to Esther).

5. Current hinder:
As MULTIPLY we can make use of current hinder to show how we can contribute: What can we do in order to make sure that the products the MULTIPLY Platform is producing will be

considered as an accepted standard by the community? We should take the current hinder as listed on board, since then we would have added value.

For the first hinder, not being aware of Sentinel derived products), we can show that we will provide tools directly aimed at Sentinel 3. This is quite new but in a few years there will be more papers out on S3, while we still start. This implies that our proof of concept should be released soon. We want to publish a paper in 2017, so what is needed to realise this? What is the content of the paper: vision, opportunities, challenges? The audience should be the user community so you would need a full blown example (on a good demonstrator case or e.g. for entire Europe) and it should be a research oriented paper on the subject of showing the added value of the MULTIPLY Platform. The approach we take is: decide when we want the paper to be published and work back on what to do when to realise the publication in that planned period. The paper will be discussed in more detail during the next telcon (Peter/Jose).

Uncertainty: we should stress over and over again that the MULTIPLY Platform deals with uncertainty and that it delivers gap free data products,

Lack of reference data could be solved with providing a meta data catalogue at the end of the project. This can't be solved right now.

A consistent atmospheric correction is an integral part of the work to be done.

6. Why not use MULTIPLY?

The first one mentioned (more diverse and internally consistent variables) is something MULTIPLY offers. Show expertise and make sure the MULTIPLY Platform is properly designed for a particular type of data. MULTIPLY tries to combine multiple sensors. This is currently not done because other platforms have no framework to do this consistently. This is where MULTIPLY comes in.

One of the key problems faced by many users is that they want to use a particular product but they don't know the compatibility of the different platforms. The MULTIPLY approach is to combine things that now are separate. Although at the moment we call it a Platform (of which there are many), maybe we need to rephrase this. How to do this, needs some thought and time. However, on the website we need to make sure that we explain why we differ from other existing platforms and why the MULTIPLY Platform is good. Or we need to try to become part of a bigger platform and integrate outcomes within other platforms.

We will not be able to solve all requirements, but we need to clearly communicate (e.g. through our website) around the functionality we do offer.

7. How can MP beat the product you are using?

When it comes to downloading data we need to make sure to have the platform where a mirror of Sentinel data exists to ensure the required speed.
There are no comparable services at the moment. Although there was a toolbox a couple of years back this offered partial solutions that we want to deal with as a whole.

We need to stress out that we go beyond FaPar and LAI

At the moment it is not clear what GeoMultiSens offers since the website is not very informative. It seems to be a platform that integrates data from multiple sources but the interface they use is not known. GeoMultiSens maybe a user or a competitor. Since this is not clear at this point it will be worth to look into this.

8. Any additional remarks?

- All participants are encouraged to send the questionnaire to their contacts (cc Esther) in order to get as much input from different types of users and stakeholders as possible.
- Traceability is important (no black box).
- At the Midterm meeting we will try to invite more users and have a stronger discussion.
- To 'sell' MULTIPLY to the community aimed a paper is the best way. People like to have a paper coming with a product.
- A competing approach is to just run LAI (done by another group from UCL) on the same sites we are aiming at. José could get in touch with them.
- Guido: the multi sensor aspect of MULTIPLY is a crucial added value. EnMAP Hyperspectral may also be an interesting tool to include.
- Keep eyes open to see what is out there and pull all together.
- If you can demonstrate consistency between Sentinel1, 2 and 3 everyone will love that.
- Make a segmentation of different user groups, but be very specific. Take that into consideration.

9. Next telcon

The next Telcon will be at 9 September 2016, 10h00-11h00(CEST). Alex will not be able to join.

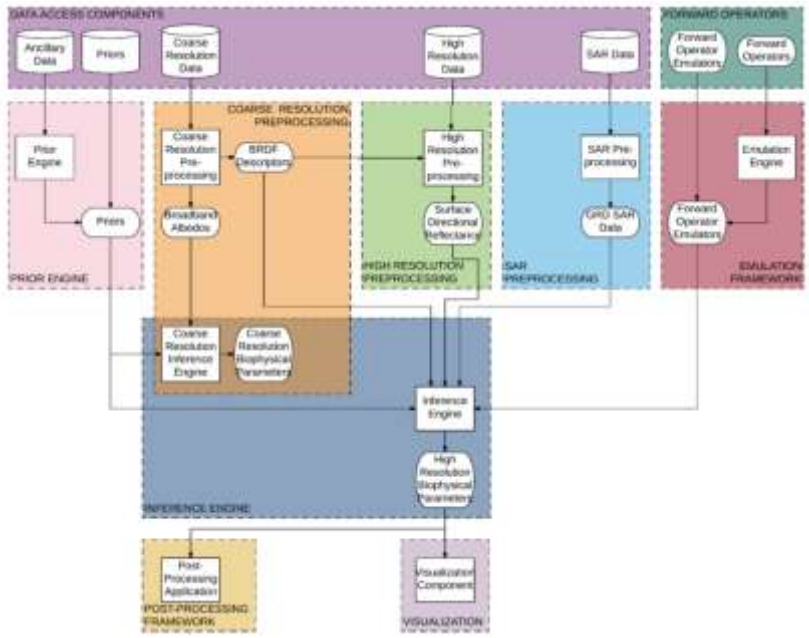
MULTIPLY Consortium Meeting 8 February 2017

Participants: Alex Löw (LMU), Thomas Weiss (LMU), Thomas Ramsauer (LMU), José Gomez Dans (UCL), Jon Styles (Assimila), Nicola Pounder (Assimila), Carsten Brockmann (BC), Tonio Fincke (BC), Lea Hallik (TO), Daniel Kindred (ADAS), Peter van Bodegom (UL), Esther Philips (UL)
Absent: Philippe Peylin (LSCE UVS), Emilio (UAH)

Calculations: what time improvement is required?
Within the Synergy project, which is using Sentinel 2 & 3, discussions are ongoing on how to speed up the calculations. MULTIPLY should evolve based on the results of Synergy. Within Synergy a Kalman filter approach will be implemented to reduce the number of iterations from 100s to 2-3 iterations. Moreover the (huge) uncertainty matrix will be split up in small matrices of 10x10 across the diagonal, each of which is solved independently and can be solved in parallel. UCL aims to implement the calculation process in the Synergy project and then start working on smoothening this for the MULTIPLY Platform, e.g. by additionally incorporating a Kalman smoother to decrease the initially high uncertainties. At least, it seems as if we can keep the promise to deliver 'near real time'. José will make calculations to see how much faster MULTIPLY will become and still needs to be.

The prototype should work very soon. Once this is the case, UCL and BC will discuss the best way to implement. The Kalman filter is going to be implemented this month, but the Kalman smoother not yet.

MULTIPLY Platform
The new version of EOLDAS (see above) will be ready in 2 months' time. The new version should be divided into different modules in order to test these separately. The Github site should reflect this modular structure. For development and testing, a local copy can be made. However, once a module works, it should be uploaded in Github as should every newly tested version of a module. Interfaces between the modules must also be uploaded in Github. Acting this way also provides for better version management. The overview of the various modules (below) suggests which pieces of software are needed. To each module we still need to add who will deliver these.



All connection lines mean “it is used by”.
All rectangles at the end represent a piece of software.
Some modifications are still needed for the scheme.

The idea is posed to clarify the pre-processing boxes in more detail, e.g. atmospheric correction. Hence, each of the data access components also have a module to retrieve these data sources and potentially to harmonize the data. The corresponding rectangular boxes should be added to indicate the additional pieces of software needed and the people responsible for those. BC explains the schedule is a conceptual way of presenting.

Codes for data retrieval to put on Github:
UCL: Landsat, Sentinel
LMU: Sentinel 1
BC: Sentinel 2 & 3

LMU is involved in the SAR Pre-Processing. The pre-processing steps are implemented in a snap toolbox basically. Thomas Weiss is using SNAP API, which would be a sub-Github project. Thomas will upload the script for the process and the sub-process.

People involved in the Prior Engine are Amie Corbin (UL) and Thomas Ramsauer (LMU). Nothing is scripted yet. There can be one or two separate prior engines. The interface is important.

Note that the Coarse Resolution Pre-Processing goes into the Coarse Resolution Inference Engine, which is (to major extents) the same code as the high resolution Inference Engine. The Inference Engine is basically a function with the name e.g. “read prior”. The people involved in and responsible for these blocks are Tonio Fincke (BC) and José Gomez Dans (UCL), with support of Nicola Pounder (Assimila).

Responsible for the Emulation Framework is UCL. The Framework generates emulators and the part that is reading and executing these. The Emulation Engine is available on Github. The output is a datafile that is used by the Inference Engine. It needs to be run and stored and version control is important. At the moment the datafiles are binary. Statistics are needed to know which emulator was run by which version.

The last 2 boxes relate to post-processing and are the responsibility of UL and others and may have to be split in different modules because it is composed of different components (e.g. disturbance, biodiversity).

Alex: baseline assumption High Resolution and SAR Pre-Processing are already on common grid. These boxes will come together. Because this may look like a single file, how does the Inference Engine deal with the outcomes? Do we need an extra box on Input Data Harmonisation Pre-Processing? For Sentinel 1 and 2 it is decided to assume that they work on the same grid and with the same projection.Carsten: If the Inference Engine needs a certain set of inputs this can be configured with rules. This could be a tool BC can use to make things speak to each other. Carsten will have to check.

To make Github work and to go towards the beta version, the following has been decided:
A Github Multiply project will be initiated fully based on dummy codes and dummy variables (or synthetic data) in order to define and test the various interfaces (defining inputs and outputs). Tonio is responsible for this. In the meantime, the Synergy project will be finished. After the Synergy project has finished, the dummy version will be replaced by the Synergy version, turned into the proposed modular structure. It is noted that Synergy mostly simply refers to the Inference Engine.
From then onwards, people can improve and replace individual modules. If one module is updated, it must still be able to 'talk' to the rest of the modules in the system we use. This is the responsibility of the programmer of the respective module. After replacement of the module, the version management is being called upon.

Actions to take:

- Drafting an Input-Output definition document could be useful to know what is expected as input and what is supposed to be output (action: all consortium members)
- An internal data structure to hold data should be set-up (action: BC).
- Raw products need to be entered into an internal consistent format and structure, the requirements for which need to be formulated. (action: all consortium members)
- Diverse input must be harmonised in a common data structure, quite close to the raw data.
- A systematic way of storing all data must be taken care of (action: BC).
- BC monitors the process and collects data component codes.
- Time of implementation to Github: end of March
- Corresponding data will be uploaded by ftp.

"From that time onwards we will be working on this repository", the repository being version 0 that will be replaced by updates along the project.

Strategy for testing:

- Define what the tests of combined process/products look like.

- Make sure the tests run automatically.
- Each of the white boxes will have unit tests implemented.
- For coloured boxes tests will be run on how they work together.
- If we know the interfaces, these tests could be defined already (even with dummies).
- The first tests will be done with a dummy.
- If it works, the dummy will be exchanged.
- If a module is not working, the person who placed it has to solve the problem.
- The partner who develops will do the testing.
- Together integration tests with the dummies will be developed.

By the end of March everything should be available, after which the testing will be done. The resulting milestone is a combined test of the MULTIPLY Platform.

Demonstrator and validation sites
The list of validation sites (D5.2) should be updated and a decision should be taken on:

- which sites are most important?
- which sampling campaigns are most important?
- Where can additional sites be set-up?
- What are the corner coordinates of each relevant site

Because no single satellite is providing all the required information, data sampling is needed. Campaigns to collect data should be planned based on what the partner thinks is the best period.

Alex is not sure on the Science hub data archive. Data should be downloaded, but this archive may disappear. Therefore it is decided that one institution is responsible to safeguard the data of one sensor or dataprovider:
BC: Sentinel 3
UCL: Modis & Landsat 8, Sentinel 2
LMU: Sentinel 1
For the overall system we need some kind of metadata catalogue. The system must recognise if data is available or needs to be downloaded.

Peter will talk to Luis Moreno and will revisit the ENMAP catalogue. This catalogue indicates all data sources and they have open policy.

ADAS can contribute results over the last 2 years. Do we ‘keep things in the bank’ and split up between this year and next year? What would we lose by collecting the validation data now? Satellites are flying now, we don’t know what will happen in 2 years. Daniel is interested in year to year variation as well.

BC mentions two online examples:
http://brockmann-consult.de/glass-granules/catalog/search
http://www.brockmann-consult.de/calvalus/

Peter will coordinate the table that also needs to contain locations and satellites. It will be discussed during the next telcon. Decisions are then also needed on the required parameters to measure on the ground and the frequency to visit.

Demonstrator sites:

1. A number of fluxnet sites need to be selected (for satellite cross-validation)
2. For the crop demonstrator we start with East England and expand it later to include e.g. the Netherlands
3. For disturbances, Angola will be selected
4. For the DGVM, Europe is selected. However, for this demonstrator only coarse resolution data will be used.

. BC and UCL will discuss about how to batch things and how to make it available.

Impact

A direct result of MULTIPLY is job creation, because of the PhD Students that are appointed to the project. However, it is important to be aware of opportunities for activities that have a societal impact.
(e.g. press releases, articles in popular magazines, news items). All activities in the area of impact or PR should be added to the Excel reporting sheet, in order to be aware of all the small and bigger activities we undertake.

Management issues

The Advisory Board will receive the User Requirement Document and the presentation that was given to the project officer and the external reviewer. They will be informed on the review meeting and on the fact that all deliverables will be published on the website after approval by the European Commission.

In June or July 2017 the next project meeting will take place at UCL. Esther will arrange a doodle for the date.
In early February 2018 the beta version will be demonstrated. At this meeting, the Advisory Board will be invited, as well as other stakeholders. The project officer intends to use this meeting for the next review. The location most probably will be Ispra.

Deliverables: a beta version should be delivered every 6 months (M18, M24)
The next 2 months v0.2 (with dummies) will be tested
The next step is to update the Inference Engine (synergy code)
In May: more than dummies, leading to the deliverable.
Then there is half a year to update the modules before the beta version is public.

Minutes Consortium Meeting d.d. 26 June 2017 - UCL, United Kingdom

Present: Alex Löw (LMU), José Gomez-Dans (UCL), Philip Lewis (UCL), Carsten Brockmann (BC), Tonio Fincke (BC), Jon Styles (Assimila), Nicola Pounder (Assimila), Daniel Kindred (RSK ADAS), Peter van Bodegom (CML, chair)
Absent: Philippe Peylin (LSCE), Lea Hallik (TO), Emilio Chuvieco (UAH)

From Dummy to Platform Beta version
There is still some confusion on high versus low resolution. The coarse resolution is sufficient for ORCHIDEE (LAI, faPAR etc. using Pinty's approach). The coarse resolution gives an expectation of reflectance for the high resolution (to allow gap filtering) and it is good for atmospheric correction, but can also introduce artefacts – although the info is a Gaussian and therefore smooth - given heterogeneity in land use within the coarse resolution pixel. However, we may need coarse resolution data as Sentinel alone might not give enough information. Current methods of downscaling completely ignore any of these effects of non-target areas (So, some steps into that direction is already an improvement). At high resolution, a binary mask will be used indicating which pixels you are interested in to solve with SAIL (e.g. using high resolution land use maps such as CORINE at 30 m for Europe) and the non-target areas are still modelled as a BRDF (which requires a linear model- already existing- and does not need an emulator) as it will affect the coarse resolution output. The link is then through reflectance and not through biophysical par., but the link has not been implemented yet. The new employee of UCL (Feng) will be able to exactly do this. Action Lewis: prioritize the activities of Feng such that this is finished in September. We may test the performance by defining a user's case testing fields adjacent to villages, forest etc. and see how it may affect the results in terms of estimated biophysical parameters.

Spatial regularization is not yet implemented (and avoided if possible). Also, at a later stage, the Kalman filter (which only moves forward in time) will be replaced by a smoother filter (which moves forward and backward). For the beta-version, KAFKA will be incorporated for the interference engine.

One way to speed up things is to run coarse resolution separately and, instead of getting the images, the intermediate products are retrieved. Ideally that demands a long-term service for coarse resolution. For instance, you can have two executables for the coarse resolution, one running the inverse RTM and the other take MCD43. MCD43 (the modis product) can be default for the beta-version. Action Jose: Ensure MODIS is coupled to the beta version.

UCL atmospheric correction (based on an emulator) is different from the one of Brockmann.
For high resolution: land cover CCI pre-processing;
For low resolution pre-processing: coarse resolution is Globalbedo-qa4ecv chain.
Will the difference lead to inconsistencies? Both should give a clear sky reflectance with uncertainty (cloud corrections are in the uncertainty and that may be a different uncertainty for each product). Part of the pre-processing is also to place all data in grids, geo-code the frame and put it in tiles (for parallelization) and keep that inside each of the boxes to allow working with multiple sensors in a box.

Technically, we should aim to have the SAR and optical running together, ie. assuming LAI and canopy water are the same state variable.

At the end of the project, you may have a validation database as standard dataset for which the changes in the system can be tested against. This may also test different ways of doing the pre-processing. Currently, we don't know whether it matters or whether being consistent would make things better or worse. So, it should be tested with the validation datasets. So, the system tests are in fact our validation exercises.

How to organize?
We need to fill the dummy with the real interface. That is the system we need to fill with content and then test, but the IODD & ICD information is needed. We will use the dummy to fill it based on user case 1, then see how much you need to modify it additionally with user case 2 etc.. We will also use the 3 cases to define the generality/functionality of the beta-version (to define the core of the system). In fact, we will use the user cases as system tests for the beta-version. Once all interfaces have been defined satisfactorily, an API can be provided to manage the interface.

Going through the various interfaces:
Prior: The Land cover CCI should always be part of the data access to link it to a LUT of priors AND for masking ''non-vegetation stuff'' at high resolution. By default we provide a time-space land cover and LUT. If the user wants, he/she replaces that by his/her own time-space land cover and LUT (user's job is to prefill the space-time of the biophysical state) in the data access components. The data access component also needs to connect to the user interface to define space-time state etc. (the orchestrator should talk with the user). User interface connects to the front-end only (i.e. to data access). The expert user does not need an interface.
The prior state will depend on the model. So, if the user changes the model, then a different state for the prior needs to be filled. The forward operator from the emulator defines the state for which state variables are needed. This is then communicated to the prior engine to be pre-filled and can be done directly in the prior engine once the user defined the problem, which internally should be translated into a required state).

The prior engine would ideally also include a simple model for senescence, as that strongly improves the LAI retrieval. The increase in senescence is fast and may be predicted as f(T, light, drought) with species-specific parameters: Action Amie.

The interfaces of the other components were also discussed and modifications to the interface documents were made (see separate documentation).

Alex and Jose will prepare a 1 A4 on the various modes in which the platform may be run (primarily whether one defined time-space block vs ongoing time series analysis) and indicate the differences/implications for using/defining the platform including interfaces, filter, smoother vs. variational and practical implications for the processing chain.

Coarse resolution preprocessing still needs to be better separated from the inference engine. Action: Jon/Nicola

Action: Jose needs to better specify the different cases of resources of the inference engine.

The main box missing is the user interface. User interface for the beta version is a GUI, simply a web form (Brockmann to develop).

We will provide basic visualization of the outputs, maps etc. and use the workshop to find out what users want. The visualization will be done afterwards to explain the cutting edge aspects of the platform, like the entropy and uncertainties (show examples on the kind of things they can do). See also: http://apps.sentinel-hub.com/sentinel-playground/ for inspiration on what might be possible.

Tonio will take the interface documents and convert them to Github interfaces and from there on start the iterations to get it working. Tonio will also arrange that more partners get administrative right on Github. Dummies should be replaced by actual code by the end of September.

Beta-version workshop

We will document the dummy performance (once that is really doing something) in parallel while developing (by Alex with Tonio, Alex to be made administrator for this) and combine that to iphython notebooks e.g. with exercises.
Most of the training material is not using the engine, but explaining how the engine works. We aim for about 30 people, expert users (groups leaders, postdocs), people with hands-on expertise.

Aims of the workshop:

1. Make an inventory of whether the beta-version fulfils the requirements of key users.
2. Opportunity for users to provide feedback such that the product can be further improved in contents and presentation;
    a. Influence the process -> we need to make that an integral part of the programme;
    b. invite people to give their opinions.
3. How to engage them/make them come?
    a. Based on the demonstration, we can help them getting started with the beta-version after the workshop (that may be the trigger for them to come);
    b. how does the platform help the users;
    c. giving support afterwards.
4. Building an active user community after the workshop. (how to optimize the platform such that it can be used directly by them afterwards).

Whom to invite? A list will be uploaded in the VRE. Please, add names, contact information and their expertise/interest to explain their invitation.
For the demonstration case EnMAP we may run the Munich site with Multiply and the Enmap machinery as demonstrator (invite Tobi, LMU to the workshop).
Maybe the workshop should be held not at ESA-itself but in a location on the top of the hill at Frascati to have the group together. Action Lewis: Provide the name of that place to Peter/Esther.

Day 1 (start at lunch): concepts of the engine, without using the engine.
Day 2: We will present the four cases to play with.
Day 3: Evaluate the experiences on how testers may use it for their own case (ending at noon).
Day 4: Consortium meeting
For day 2, we need to host the datasets somewhere (e.g. run them on CEMS). The biggest time consumer is downloading the data. So, that needs to be prepared: Action Jose.

Other issues

- The partners are worried about the absence of Alcala. So far, they are not involved nor showing any interest in the project. Peter will contact them to discuss their involvement and urge them to be actively present at the next meeting.
- Multiply will be presented at the open science conference next year, not this year.
- The access rights need to be arranged. We need to discuss how to organize the work flow. We need a master branch and a branch to work on. You need for each module someone in charge of it (master user) and someone on top (Tonio). Tonio will arrange this.

Hosting:

CEMS, CODE-DE (German server) also offers a service for free till 2019: An announcement of opportunity is published that gives the option to experiment (https://code-de.org/en/news). This would be good for the middle-long term for Multiply. Alex and Carsten will follow up on this (not much competition probably with other people, given that they start up now, they might be supportive). CEMS should be contacted for hosting the datasets for the workshop: Jose to find out. For demonstrator it is good to have different systems. The German Cloud Server is https://code-de.org/.

MULTIPLY Consortium Meeting 5 February 2018

Participants:
Gerardo Lopez, Jose Gomez, Nicola Pounder, Daniel Kindred, Leon Hauser, Amie Corbin, Thomas Ramsauer, Thomas Weiss, Philip Marzahn, Carsten Brockmann, Lucretia Petinari, Tonio Fincke, Feng Yin, Joris Timmermans, Lea Hallik, Philippe Peylin, Peter van Bodegom, Esther Philips (minutes)

Agenda:

1. Workshop: tasks and roles
2. What to communicate?

Programme Tuesday, 6 February 2018

1. Presentation on the justification of the MULTIPLY Platform: Peter will introduce (presentation will be send for last comments).

2. Presentation on the structure and components of the MULTIPLY Platform: Carsten and Tonio will present (presentation will be send for last comments)

3. Interactive demonstration and do-it-yourself assignments of concepts on the fore front of satellite data retrieval: coarse resolution (Jose): demonstrating coarse resolution change. Presentation of components. Jose will work on the demo. Why are we doing this this way? (kind of philosophy)

4. Continuation of interactive demonstration and do-it-yourself assignments of concepts on the forefront of satellite data retrieval, pre-processing:
   Coarse resolution preprocessing: Geraldo
   Feng Yin (20 slides + demo). Show that it is a problem solved. Details can be discussed during the poster session for interested people.
   Thomas Weiss: click some buttons during interactive part to show the difference between running preprocessing on full data versus a small part. Duration of preprocessing depends on area.
   Findings retrieved during work on pre-processing to get things running and get a retrieval as a flow chart will show what is really done.
   Time: around 30 minutes.

Pitches of posters, followed by the icebreaker reception and poster presentation.

We intend to organize a workshop together with the 4 DIAS project groups. Peter hopes to discuss this with Massimo during this week.

Programme Wednesday, 7 February 2018
3 case studies where users can explore. The exercises build up.

First: follow up on Jose's presentation of Tuesday.
Play around: what do priors do to retrieval?
Teaching material on approach was put together. It was tested and it worked.

Case study 1
First: introduction.
Then: work in groups. Walk around and assist
Visualizing: 10 minutes
Exercise with predictions: familiarize people with parameters 10-15 minutes
Time series: 30 minutes
Continuing the experiment with a dynamic model: 30 minutes

Case study 2
agricultural view (central part of Spain) – optical high resolution.
Artificially increase observations.

Case study 3
SAR processing brought in as well.

To have experienced coding persons in both rooms.
Room A: Jose, Nicola, Thomas and Tonio
Room B: Joris, Geraldo, Thomas, Feng

Preparation for Wednesday:

- Case 1: nothing. All running on the notebooks that were prepared last year by Jose and Assimila.
- Case 2 + 3: check if Sentinel 2 data is properly read. Everything contained in 1 notebook. (Joris, Feng, Nicola, Geraldo)
- Physical parameters for microwave model need to be integrated. Jose, Thomas, Thomas.
- Tonio: in between.
- Virtual machines, check firewalls, all should be up and running tomorrow afternoon (Joris) (before actual exercises on Wednesday).
- Short introduction to the case studies (Joris & Jose)

Programme Thursday, 8 February 2018
Questionnaire forms the basis of the discussion.
The current state of the MULTIPLY Platform is "in between" (work in progress).
Specific things on programmatic aspects, visualization, etc. will not be added, since these questions were asked in the 1st questionnaire we distributed (2016). The questionnaire will be put in context by distilling a roadmap.

We need:

- A network of people who want to be involved.
- Feedback based on experiences and collate.
- Try to achieve some bonding to help us achieve the impact and get people interested in more extensive actions.

What else?
Thursday we will have a plenary discussion based on the answers to the questionnaires. Per room (Wednesday) a rapporteur from the participants will be asked to give a short presentation on their findings.

Thursday afternoon, the consortium will have a meeting with Massimo Ciscato.
Friday morning the internal consortium meeting takes place.

What to communicate?
We will be fully open, while focusing on what we have now. Not on what we don't have.

What do we want to stress, showing what we can do?
We already emphasize this through the case studies.

Where do we want to go after the end of the project?

1. Ideally we will have the Multiply Platform up and running in a DIAS type of system.
2. If we would focus on 1 particular Copernicus service as output of the platform, this could be a way to go (suggestion Massimo). We may choose for a land service or a climate service.
3. For the climate services, C3S might have potential in 3 years' time. Since it is a running service, there is no earlier option. C3S will get a new set up after 3 years, so there could be some possibilities then. We can at least talk about it since usually there is some momentum at some point. At least, the multiply platform produces many of the parameters that are already in the C3S system, although now those are produced separately, whereas they could be produced consistently by the multiply platform. This might again be coupled to DIAS. Land services are more on classification, although the multiply platform may assist deriving these variables if connected to proba-V.
4. Will it be an option to receive financing from ESA and the climate change program?

The software will be there so people can always download it and run it at home or where ever they want.
But:

- Do we want to maintain the MULTIPLY Platform if the service will not be paid by DIAS or a different host?
- People install and then do have a question. Where can they ask? Where can they report a bug?

The answer should be clear. You need one financing organisation that supports open source.
The MULTIPLY Platform will be a big thing to maintain. That is why Massimo suggested to connect to an existing (Copernicus) service. However, Copernicus pays for a product, not for software.

How to continue beyond the projects' end will be part of the discussion on Thursday morning.
Our main target will be to make the MULTIPLY Platform attractive to be used in Copernicus climate service and land service (local component, Pan European component and global component).
Our main selling point is the combination of coarse and high resolution in one package and being able to run that global.
Another big selling point is that the MULTIPLY Platform deals with uncertainty.
These are our main targets. We are not sure how to achieve it but we have time.
Beyond this we will provide the software as the virtual machine.
But on the long term funding is needed to maintain it and we need to dialogue with users on how to realise this. What would users be willing to contribute to maintenance?

Philippe Peylin indicates that for him having coherence between different variables, based on the same observations as part of the model, is the most important selling point. For this we should consider to present MULTIPLY at the (annual) meeting of climate modellers where preparations for observations are made and discussed.

Another point of attention is that we should put more emphasis on dissemination and (societal) impact.

Roadmap MULTIPLY

Short term developments towards release of beta-version on github end March:

1. Clean-up of the inference engine and basic documentation: Jose – deadline 1 March
2. Consolidation of the interface done with weekly virtual coding workshop: Tonio with Jose, Gerardo, Nicola, Joris. Deadline for consolidation: early March
3. Check license: Brockmann and Assimila decide. Deadline early March
4. Orchestrator development: starting with answering to open questions, most work after 1 March: finish end March
5. Modify relevant parts in other modules, including bits coming from Kafka. Deadline end March: All coders
6. Document all modules and specifically document the flow chart and work flow from A to Z as well as the variables list. Deadline end March. Coordinated by Joris
7. Set-up e-mail for helpdesk function: Esther

After release and before 30 June:

1. Release to test users
2. Ingest the functional diversity module into the Multiply Platform: Leon
3. Determine consistent propagation of uncertainties at high resolution based on (provided) input uncertainties: Feng.
4. Prototype of coupled optical and microwave to be made ready: Thomas, Philip, Joris, Jose
5. Prototype of analysis of time series with error propagation
6. Prototype of SAR retrieval beyond the c-band.
7. Allows pre-processed Proba-V data to be ingested in the coarse-resolution engine.

After 30 June but before DIAS presentation (coding workshops to be organized to make it work…):

1. Prototype of coupling high and coarse resolution.
2. Use existing cloud masks
3. Use pixel-ID field to identify non-vegetated pixels.
4. Potentially to be postponed: dynamically coupling the inference engine to priors

Minutes MULTIPLY meeting with Massimo Ciscato
08-02-2018

Massimo:
It was an interesting workshop.
The official project review will take place at the next meeting. Deliverables are due at the end of June 2018. The project report (technical & financial) are also due on 30 June + 60 days. A brief report on status of the project was provided.
Before the review meeting the draft technical report should be available. This should not be a problem. The financial report will take longer. The meeting will be scheduled around the 3rd week of July. Another option is the end of August. Massimo will check when the reviewer is available. The meeting was planned in Tartu but it could also be hosted in Brussels.
The recently submitted report will not be formerly approved. Approval of these and the deliverables will be after the next review meeting.

Management perspective:

- Amendments are taken care of and have been approved by the EU.
- The next action is to adjust responsibilities. Assimila and UCL will have to inform the EU on how to divide the tasks. *(Note from Esther: this has been done and the DoA was updated when the amendment was submitted)*
- UCL needs some extra manpower since Lewis is ill. For this an overview of the available budget is needed. Esther will contact Maria (financial responsible at UCL) on the budget shift.
- The change for Tartu is being dealt with.
- Where UAH is concerned, it is important to decide on their contribution. What are the most useful tasks for them to provide will be discussed tomorrow morning at the Consortium Meeting.
- After the Brussels meeting with Sensagri and ECOLass we were looking for synergies. Unfortunately we have not heard from Sensagri (although Francesco from Sensagri was the workshop and some aspects of Sensagri were discussed). Massimo will get in contact with the coordinator of Sensagri to request the presentation.
- A follow-up of the Brussels meeting is planned and ideally EEA, JRC and EC should be present in such next meeting. Massimo will try to set it up.

For the EU it is important that the project delivers what is promised. Small transfers are allowed without amendments.

The technical WPs are all under control.

Other points of discussion:
Exploitation
There is still time to think about the exploitation upon the end of the project.
We already discussed the option of DIAS and are already in contact with some if the other projects (Ecolass). We will stay in contact and follow up on the idea of Linda.
There are DIAS projects working on developing intermediate tools. Massimo will put us in contact. Several of the DIAS providers are also involved in these kind of projects. Maybe MULTIPLY could be a showcase and test-user to a more exploration kind of project in DIAS.
The MULTIPLY platform could serve as a DIAS platform to allow deriving EO-products

within DIAS. So, there should be an interest from DIAS operators to get MULTIPLY on board. We will explore the possibilities for exploitation of the MULTIPLY platform with them.

Milestones due 2nd reporting period							
Nº	**Title**	**Lead**	**Due M**	**Description**	**WP No.**	**Status**	**Activities**
M2	User requirements workshop	UL	12	User workshop held and report available (D9.2)	**6, 8, 9 ...**	☑	This has been dealt with in July 2016.
M3	Final selection of land surface state variables	UL	12	Summary published in first report to Commission (D1.3)	**2, 3, 4 ...**	☑	A summary was published in the 1st technical report to the European Commission.
M5	Final selection of validation sites	UCL	12	Status report (D5.2)	**5**	☑	Several field campaigns in different locations (the Netherlands, Finland, Estonia, United Kingdom, Germany) have been carried out to collect field measurements of traits, which will be used in the validation exercises. The selection of the validation sites has been finalised.
M6	Design of pre-processing chain	BC	24	Design follows System Specification Document: Chapter Data Access Components (D6.4)	**4, 6, 8 ...**	☑	A detailed documentation of the pre-processing chain for high resolution optical (UCL) and microwave data (LMU) as well as a pre-processing framework of ancillary data for prior production (LMU, UL) was provided (D4.2)
M7	Forward operator codes available as open-source codes	LMU	24	Status report (D2.1, D2.2) and on-line availability of code	**2, 6**	☑	Optical (UCL) and microwave (LMU) forward operator codes were defined, implemented in the beta version of the MULTIPLY platform and stored in a GitHub repository of the MULTIPLY project (D2.2)
M8	WebGIS interface for input and output data	BC	24	Design follows System Specification Document: Chapter Input Data Exploration Tools (D8.2)	6, 8		In our proposal, we had indicated to make a Web GIS interface. Following the key users requirements workshop, it became clear that users do not prefer a web GIS interface. It seems our key users split into two groups. One group prefers to be able to modify all code themselves. For that group, a good Github depository with

							libraries and documentation is the best way to promote use. Another group of users want to be able to get relevant and high quality data products for their regions of interest. For that group, a visualisation tool (but not with a web GIS interface) will be made. This tool should provide data access, selection of region of interest, provision of the data product and visualisation of the data product. For this milestone, we will have the design on what the interface would look like. This was changed in the deliverable, but not in the milestones. We will discuss this further during the review meeting in June. report on whether to do or not and why.

Minutes Consortium Meeting d.d. 9 February 2018

Present: Philip Marzahn (LMU), Thomas Weiss (LMU), Thomas Ramsauer (LMU), José Gomez-Dans (UCL), Feng Yin (UCL), Carsten Brockmann (BC), Tonio Fincke (BC), Lea Hallik (TO), Javier Salas Rey (UAH), Lucrecia Pettinari (UAH), Nicola Pounder (Assimila), Gerardo Lopez Saldana (Assimila), Daniel Kindred (ADAS), Peter van Bodegom (CML, chair), Amie Corbin (CML), Joris Timmermans (CML), Leon Hauser (CML), Esther Philips (CML, minutes)

1) **Lessons learned:**
   a) **Documentation must be made**
   b) **The real beta-version needs to be released within 2 months**

   Discussion on modularity

   i) Tweak beta version (Jose; deadline 26 February 2018)

   ii) Set interfaces in stone (deadline 26 February 2018).

   Within 1 month, an overview of all components and workflow within the GitHub organization must be available (**Tonio**). This should help creating a logical understanding on how GitHub is organized, also to external stakeholders.

   Discussion:

   Tonio: we need something to document on GitHub. The users wanted to get something from GitHub with which they can work, e.g. a repository with more detailed info or documentation on the inference engine. The total interface (how to get the platform to run) is another thing.

   Jose: we must make sure to have some documentation on individual modules. Even more so, we must think about the functioning of the individual modules and their connections to the whole.

   Tonio: the orchestrator would pull everything together but it doesn't work yet.

   Peter: how much time does it take to make the orchestrator work?

   Tonio: to get it to work in the form we have it now, means the need to integrate the inference engine and to have the interfaces defined. This is a joint effort. How much time it takes therefore depends on when we can spend some time to see how it should work. If support from others is given now, it should be up and running in February.

   Carsten: February is optimistic. We must first agree on interfaces and fix those. Consolidation of different modules and agreeing and fixing of the interfaces definition can be done simultaneously. When this has finished, the orchestrator can be fixed.

   Joris: the interfaces were defined during the consortium meeting at Brockmann. So, are these definitions different to what needs to be decided upon? And how are these different?

   Peter: interfaces have been defined, but current codes do not use those interfaces as defined. Does this imply that the interfaces are wrong? Or should the code be moved to compile to the interfaces as defined.

   Jose: the interfaces were not yet tested, so obviously they would change. It should be minor changes only. Also parts need to be moved to other modules, e.g. the notebook using Sentinel 2 data should be updated. Now it is on the inference engine as Python script.

Gerardo: revisit interfaces. We want flexibility (e.g. pointers to input files provided). Then integration between components will be easier.
Tonio: people want to pick individual modules they need.
Gerardo: you would need something a bit more clever to really be able to separate things.
Peter: start by stripping Kafka and check what it is doing. Then we can automatically see what the engine should look like.
Jose: we have a clear view on what we need from the data.
Tonio: do we run a Kafka test?
Jose: when using observations, all those have same the interface. One method is used, always in the same form.
Peter: Kafka has components. These should be stripped out so that Kafka becomes an inference engine alone. As for the inference engine as such: are there things that still need to be tweaked? Or can we leave it the way it is?
Nicola: the inference engine is not working all the time and still needs some testing.
Peter: We want to release the MULTIPLY Platform GitHub to users in 1-2 months from now. It will not be perfect, but it contains components that talk to each other. It is preferred to have something complete (components & orchestrator) with constraints and limitations. In due time the inference engine will be replaced by version 1.0 with the same interface but with more functionalities.
We do need to have version 0.0 in place earlier. If we wait for tweaking up time series, it will be too long to show to users.
Jose: it is sensible to split and work on things in parallel.
Peter: What needs to be done now?
Jose: some documentation is needed, rather now than later because you can at the same time do a clean up and line up things that go to other places. Documentation is not critical but some is needed and it will help the other people working on this as well.
Nicola: it doesn't have to be detailed documentation. Writing down the basics that someone else needs to know to take a code and run it is sufficient.
Peter: when can the basic documentation & clean up be ready? We set a deadline of 1 March. Should the discussion on interface also take place? Or in parallel or earlier?
Tonio: this should be done in parallel and we must try to avoid unnecessary work.
Peter: how can we connect the clean up with consolidation of the interface?
Tonio: this is the same process, so communication is needed.
Tonio has some questions about what might be done to the inference engine to be able to talk to other modules.
Peter: how will this be organised?
Tonio: the optimal way is to organise a workshop with the coding group.
Peter: it would be good to organise a virtual coding workshop on a weekly basis at a reserved timeslot and day to make sure it will run and continue. We need to have things ready in 1 to 2 months.

Peter: what needs to be done in other modules? Many things will be affected if KAFKA will be stripped. We need to create a roadmap on what needs to be done at what time. One roadmap should focus on the short term and one on the longer term (next 8 -9 months). Short term parts coming out of Kafka should be incorporated in other components. Are there other things that need to be done in other components to release something?

Thomas: does the prior engine need to get information from the inference engine to make updates?
Peter: this will be part of the development from version 0 to 1.
Tonio: but the prior engine and the inference engine need to communicate.
Peter: this is done through the orchestrator. Will it have an effect on the orchestrator? Or is there an option to include this later in the current design.
Tonio: if we do it later, we need to be aware of that in the design. When we have an orchestrator which would assign workloads on clusters this becomes very important. The current design needs to take that into account.
Peter: the prior engine can make basic runs.
Tonio: local data access works.

Peter: Considering the preprocessing modules: is there any work to be done? And for the high resolution?
Feng: working on that. The code is fine, everything is working.
Tonio: the atmospheric part needs work, e.g. pixel identification.
Feng: the current mask code works. It can help to improve it.
Peter: that should be part of the update for improvements.
Jose: still missing from the atmospheric correction is uncertainty. This is important.
Carsten: preprocessing works only with Sentinel data. Do we restrict to certain files coming from a certain provider? Is this a problem and can we solve it later?
Peter: We can solve it later. Now we need to show something to users that works from A to Z under specific conditions.
Gerardo: prior, global or not? Everything except for the leaf area index is global.
Nicola: We have to collate what we are doing.

Peter: what should we incorporate in the 1st release? Do we include the variable list now or later? Decision is to include the list now and update it later.
Joris: will make the variable list in combination with a flow chart: a graph showing variables in relation to models. If we add a model, it will add a list of variables.

Peter: is this all that needs to be done to have things released?
Gerardo: keeping one of the virtual machines might be useful for testing and for having all data at one central place instead of having some data distributed here and there.
Peter: keeping one virtual machine is fine.

Question Tonio has in relation to the orchestrator:
Both Python 2 and Python 3 are used. We need to decide on one version to use.
Decision: **Python 3**.
This means everything for which Python 2 was used, needs to be moved to Python 3.

Another question is which license we will use: GPL or MIT?
MIT would allow everyone to use it. GPL also but there software should be open source. MIT sounds more attractive.
Peter: did pre-cursors affect the choice for license?
Tonio: so far not.
Peter: pieces of software were developed by BC and UCL before MULTIPLY started.

Carsten: The license can be changed, so that is no problem.
Peter: at LMU, UL and UCL there is no particular preference.
Nicola: Assimila will check preferences and communicate on this with BC. They together will decide and the others agree.
Carsten: Wants to discuss the options with the legal department and decide after that. This will be approximately 2 weeks before the release date.

Conclusion:
We start with cleaning up and provide basic documentation on KAFKA. In parallel weekly virtual coding workshops will be organised. Based on that and in relation with questions the orchestrator starts. Flowchart A-Z describes and consolidates all that was done. End of March will be the release.

- Kafka will be stripped
- Interface consolidated
- Modules communicating
- We need documentation on how to go from A to Z. This needs to be simple, e.g. an example with comments, a bit like a recipe. **Joris** will take the lead in consultation with Tonio and Gerardo.
- Complete all Python libraries (**all module managers**)
- Variable list to be prepared (**Joris & José**)
- Basic documentation for each library within 1 month (**all module managers**)

Deadline: Monday, 5 March 2018

c) Need for test-users: they need beta version and helpdesk.

d) When and who will develop case studies representative for users showing the entire chain?

Test users start once we release something. The test users so far are:
- Alexandra Wolanin (insights on EnMap),
- Jonathan Leon-Tavares (data, Sentinel 2 data),
- Caglar Kucuk

Joris: Will they use the whole system or the part they need?
Peter: Ideally those 3 test if they can work with what we release. Also related with the helpdesk function. We need some function for that, at least for these test users.
Carsten: the email address to pose questions is coupled to GitHub and will be distributed to the team and to the test users. As soon as questions come in, someone will answer. This answer will be copied to the person monitoring if questions are followed upon. Esther will monitor and she will make sure there is a deputy in place.

e) When will we be ready for the 6 to 9 Sentinel tiles of Ecolass?

f) **Esther** will distribute the participant list, presentations and photos. If anyone has documents/photos to share, please send these to her.

g) An issue tracker is automatically part of GitHub and is visible for anyone.

2) What's missing?

a) Coupling optical versus microwave:

We did not put this up as a deliverable, but it is both the central aim of the project and one of our big added values, so we need to do it. When would this be feasible? How to couple it? That is the big question. Checking indexes and calibrating one on the basis of the other is one way to do it. We can't assume however that it is robust for a whole range of conditions.
After the release of the beta version, Thomas W, Jose and Joris will sit together to discuss how to get the SAR running as well as the coupling between optical and microwave. This coupling should be ready in 8 months from now, to present it at the DIAS meeting.

b) Postpone Proba V to 2019?
There is a strong push towards including Proba V, especially in relation to the Copernicus Land Surface. Including Proba V is part of the issues to be solved when it comes to the coarse resolution. Brockmann is developing a PreProcessor that also works for Proba V, but this is all done offline.
An option is to ask VITO people as test user for Proba V processing. They have both the computer and the manpower to give it a try. That might be easier. The main issue is merging the Proba V data with other reflectance data. Otherwise, we include the standard products of the MULTIPLY Platform (and don't need any processing). This can be done when Jonathan is ready to test and requests very limited work on our side.

At the moment there is no blending of high and spatial resolution because it is not yet known how to do it. At the moment the preprocessing only runs with high resolution or coarse resolution data. An original idea has to be developed, there is no solution yet. This is one of our tasks. However, it is yet unclear when we can plan such activities.

Missing & priorities:
0. coupling optical microwave
1. Uncertainty quantification in high resolution processing chain (no proper uncertainty for SAR).
2. Time series with proper error propagation and harmonization (in parallel with 1). Will start via other project.
3 Coupling high and coarse resolution
4. Dynamic coupling inference engine to prior: demand for more dynamic prior. Provide prior data as whole time series in advance to inference engine? Needs to be discussed.
5. Cloud mask (Feng has something that may be used). Brockmann also has a cloud shadow mask. Update plan is in place. Brockmann will check if snow is in. Pixel ID as place holder for that part.
6. Non vegetated area detection

3) What's next?

a) Validation:

Which campaigns?
How much do we need in addition to get proper validation? One of the key things is getting the right metrics. Field campaigns this summer should focus on the combination of optical and microwave.
LMU will continue and the preparation phase is ongoing. If partners (**all**) want some specifics to be measured, they can send a wish list to LMU which they will try to incorporate. LMU also measures FAPAR and air & soil temperature at a forest site,

near the Alps (in a valley). Is there a possibility to measure leaf traits from the canopy?

.

Tartu will also continue. Is it possible to measure soil moisture as well? **Lea** will check at other groups if this is possible. This is mainly a question of who will go and who will manage the data. Another option is to install a button that will measure on a battery. Lea is advised to look into this and she can always contact UL, LMU or ADAS for questions or advice.

ADAS can take samples again. This is useful, particularly throughout the season. Northampton was chosen last year. ADAS will get back to the same farm this year, but Northampton is cloudier than elsewhere. Should ADAS keep it to the same area or doesn't it matter where to go in the country? It does not matter Farmers conduct their own measures anyway.

UL: **Amie** will contact Francesco Mattia to get additional sites in southern Italy.

b) Impact

- We do need to organise a next users/developers workshop and maybe can collaborate with Chris Tammer (ESA) who organises a Summerschool in 2018.

- What to do in 2019?

There are 2 groups that require 2 products:

- An intermediate platform to be realised via DIAS. Users are GitHub experts and work individually. For this group flexibility is required.
- One group is less experienced. Visualization should be targeted at that group.

A business plan is required.

4. Societal impact & PR

- Website: There is no time to properly discuss this issue, but it needs to become more vivid and for this input from all partners is needed. Serious stuff, fun stuff, pictures, and the works.
- A monthly blog should be added to the website. This means all members at some point have to write a column. **Esther** will draft a schedule for this.
- Twitter is set up and will be connected to the website. Also here input is required. For those members that already have a Twitter account, they can send messages directly. Please contact Esther to get direct access to the account.

5. Next meeting:

Year	Period/Month	Location	Type	Duration
2018	mid July (or end of August)	Tartu, Estonia	Review Meeting with EC	2 full days
2019	January	UAH, Spain	Consortium Meeting	
2019	June/July*	LSCE, France	Consortium Meeting	
2019	December	Brussels	Closing Conference	

*This meeting may be arranged in combination with a Summerschool.

Massimo Ciscato is eager to try and arrange a presentation day with DIAS projects in the autumn 2018. We have to work towards that. This implies that a coding workshop should be coupled to the Tartu meeting to incorporate prioritized elements in the Platform and be able to show something new. We should discuss this during one of the telcons.

FEEDBACK SESSION MULTIPLY MEETING FRASCATI
8 February 2018

Participants feedback from the groups

Caglar Kucuk
Based on the programme, the expectation was that the workshop would be more interactive than it actually was. It would have been of added value if more detailed information on the b platform would have been provided. More libraries were expected to be available. That would have made working through the exercises easier. In the two-stream function, a variable was missing.
Furthermore, it was hard to understand observational uncertainty parameters. Especially if you are not experienced in this matter, it was difficult to understand which is the lower and which is the upper one.
The afternoon exercises were more exciting, although it took time to merge everything.

Nuno Carvalhais
A first general impression was that within the group there was difference in the level of expertise, as well as in expectations and questions. More details were expected on e.g. what data to use, what are tools doing, what is in the system and what are the assumptions.
There also seems to exist different perspectives on how to improve the platform. Some would prefer to be able to modify and evaluate the priors, code and data exchange between user and interface in order to have more flexibility. Others may not have so much interest in going in depth into e.g. the parameters. Different types of users will want different levels of flexibility.
There was no common agreement on the meaning of uncertainties in relation to time series. The uncertainty is highly dependent on the parameters used. It would be good for the user to have the freedom to set its own prior individually.
An interesting aspect was related to transparency: what was under the hood? A discussion took place on containers, Dockers, GitHub, and the ability to explore locally without the need to use the server.

Discussion based on responses to questionnaire (*Peter van Bodegom*)
Documentation

- Overview of all functions and all options – step by step throughout the chain
- (access to) Python libraries
- Understanding of file formats
- How to input/extract data for validation, priors, ……..
- Transparency in the deployed code
- Getting a better grasp of the work environment
- Highlight the assumptions, uncertainties and limitations of models

Discussion
The consideration is when to start drafting documentation for all functions and options and how mature the MULTIPLY Platform should be for this. Since developing the Platform is work in progress, it may be a good idea to **set a roadmap**. Especially since there is quite some variety in expectations, differing from: ‘documentation containing all details’ to ‘an overview of all functions on a high level to easy access the Platform’. The roadmap should

include how to engage users and other stakeholders. To know where to start, the audience is asked what we should document to start with?

From the perspective of the consortium we should offer clear documentation on the code. This ensures that it is clear to the user what input for the code is used, e.g. what are the functions to be caught from the inference engine, etc. The documentation should be published on a website, including information about the code itself. Some work to provide this is on its way.

What would the users need from a library environment (GitHub, Python, current Notebooks), knowing it is not fully ready but ready to be played with?

If a user wants to work with a tool like this, the first requirement is to understand the tool and have documentation on the particular modules, e.g.:

- How to input and distract data?
- what data does the user have access to?
- Is it possible to conduct a study and work both local and on a server?
- If the data is on a server, can the user conduct a local study?

If the documentation on the developed tool are to be used and relied upon, its principles should be explained. This will probably demand a proper documentation on workflow through the platform: a user should have a general idea what to put in and what can be expected as output and knowing the workflow can help here.
Another point is that some of the modules should work independently of one another.

The need for documentation for the different workflows used by different users, is sort of a starting point. Users will always need detailed documentation for everything. Practical use cases, going into more or less detail, need to be described. These cases should go from A to Z and show the entire workflow.

One of the remarks is, that is should be made clear who the users are. The role of MULTIPLY should be clarified: will it be a state-of-the-art retrieval platform for a range of parameters or will it be a platform developed for the scientific community? Will the Platform be used to better understand products or will it be a tool to retrieve parameters for end users? Answering this question should help in answering the question where the focus should be considering the documentation. The end-user will probably not be the farmer.

The call text asks us to provide a tool that enables scientists to do their work more easy and to develop the tool to be used by end-users. The primary objective is thus not a click and point tool and the secondary objective assumes some expertise. Also within the scientific group there is a wide range of experience and expertise, which leads to a difference in expectation. Our first target to enable people to work with Python libraries, etc. which means we should document the relevant aspects to start with.

There are people who try to do the same as MULTIPLY tries to do. Users need to be able to rely on compatibility and it is great to see efforts to make a Platform that is valuable for the end-user. In that sense the MULTIPLY Platform should become the standard for benchmarking: the MULTIPLY Platform is the standard that everybody has to stick to in the

end. The data provider needs to stick to the MULTIPLY Platform procedure or change it towards the benchmark.
The business model to continue the MULTIPLY Platform after the project is finished, is interesting in that perspective. Will the Platform be open source? Will it be hosted on the ESA website? Who will provide the service? It will be interesting to see what the users can tap into at the end.

The Platform allows users to derive their own model. The model will run through an emulator and will be plugged in, so users can come up with their own primary product. Some users may want the MULTIPLY Platform to run the primary product. There we will provide a default set but also give flexibility to users to plug in their own model.

If the MULTIPLY Platform will be open source, it will be used. If the Platform is supported for a long time, then users are willing to invest in a learning curve. If there is no such guarantee, users will look for someone who does. The business plan is therefore very important.

For the duration of the project the MULTIPLY Platform will be open source and there is no plan to change this afterwards. It is a clear open source activity. The business model on a service is a different question and how this will turn out depends on the users. Brockmann could do it and offer it as a service, integrated in the customers workflow. The MULTIPLY Platform would remain in public domain and become a product from the user.

At the moment it is more important to state that it is and stays open source. The user community and (stakeholder) organisations should support the Platform to keep it alive after the project. Therefore we should try to integrate the Platform in the existing community.

We will try to link the MULTIPLY Platform to DIAS, but in the end we still need a business model. Support and updates will never be for free. We are curious to hear the participants ideas on this. What kind of business models would you be willing to use or not to use? These insights would help us to frame and we still have 2 years to develop the business model.

DIAS started after the MULTIPLY project started. In DIAS, a number of the bullet points (see below) would be solved.

Making it work

- How to get real time access to data
- Offer the choice to parallelize processes
- High computer power possibility to link to diverse dataset
- Guarantee 10 years of stability (to allow scientific reproducibility)
- Keep it modular and flexible
- Promote modular and local usage and development of methods
- Provide cloud computing for the retrieval of land surface parameters
- Ensure the usage of non-public but web-based usage
- Expand to other languages like R, Matlab
- Developing a business model for services

DIAS counts 5 projects of which 4 are a commercial platform. Copernicus data would be for free.
A business model is required to provide services in this infrastructure. The MULTIPLY Platform could be provided as an intermediate-level service within DIAS: can others use the developed platform for use by their own customers or to provide a direct service to their customers? Where to position the MULTIPLY Platform: intermediate (talk to providers) or directly in the market with customers? Here we should make up our mind.

This depends on the user. Would users be willing to pay for the service to build/improve their own products?

The Ecolass project thinks scientific developers will not pay for a Platform.
However, scientists do pay for running high performance computing systems such as DIAS.

The definition of 'the scientific community' is important. Do we mean anyone who exploits the tool for his/her own product? If you use the tool for your own work, you are a scientific user. Scientific users are academics. However, the real opportunity of the service is that the community can be expanded. Tools should be accessible to people who haven't used it so far and don't yet know what possibilities are offered. We can expand to these users.

What preferences would such a more down-stream community have?
An interactive GUI to start with.
If the MULTIPLY Platform is first used at a high level, a user wants to know how it works at that level. Once the user goes further, he/she will need to know what parameters can be changed and more documentation is required. If you start with a case that allows you to click and go and allows going through all modules, a user can become more independent and more familiar.
All information, queries and results should be kept in a library so users can learn from each others experiences.

This is all good, but a choice should then be made between a virtual machine or a platform. These preferences suggest a platform. Then we need one central host to incorporate everything to indeed be able to learn from earlier experiences.
If the strategy will be to have a fully traceable system, you basically need to be able to reproduce each data product you get. Each file you get and want to replicate comes with a recipe that can be followed. This has not been given much thought yet, but is a critical part of data management: store the file, and the metadata on that file.

Ecolass is willing to test some of the parameters that come out of MULTIPLY. A large size is needed in that case; huge pack of data (9 or 10 sentinel tiles at once) and most importantly: a well-developed database of priors.
To make it work, two aspects are needed: (1) Traceability, not a black box. (2) Reliably running on a large scale. The time to run it on larger scales and longer time series should be around a week.
A global scale is a larger scale and some participants would like the time to get results to be 1 night.

A part of the question is ignored: where do you want it to happen? The infrastructure is important.

A test is run on a smaller scale to be sure. To run it on the global scale is fine if would take a week, but with the knowledge that it will not crash and the results will be what you expect. But testing is important in that case.

Demonstration of accuracy

- Validation results of biophysical parameters
- Proof that it can be used for long time series (and large areas)
- Show the uncertainties in the outcomes
- Joint evaluation/validation case studies

Visualisation

- A way to save and easily retrieve results
- Easier user interface
- Make multiply products easy to (bulk) download
- Make the uncertainties as easy to download and plots as the main products

New functionality: coupling sensors

- Full integration of high and coarse resolution to be demonstrated
- More flexibility in combining different data streams
- Integration of more satellites to increase temporal and spatial resolution
- Include data from medium resolution missions such as Proba-V
- The combined use of optical and microwave would provide an important product

New functionality: additional parameters

- Retrieve parameters which are easy to interpret for non-experts e.g. evapotranspiration
- Allow deriving biomass or parameters that can be translated to biomass
- Add canopy temperature

New functionality other

- Provide better constrained prior distributions
- Better provision of uncertainties (which uncertainty and how does this together contribute to final uncertainty?)

Communication

- Access to the platform and its documentation
- Help with potential issues that could arise when running the platform (or components thereof, e.g. emulation)
- Maintain something like a helpdesk of active user/maintainer groups
- Continue use/developers workshops
- Maintain contact/user forum for information on progress
- Distribute contact list for subject areas
- Create a mailing list with updates/releases related to the platform

We will provide access to Python libraries on GitHub but we need 1 or 2 months. We need to think about how to document things properly, show the workflow, show how things are interconnected and provide a brief description. This will not be extensive documentation. Based on priorities we will work on that to keep you on board and we hope you will start testing and giving feedback so we can improve the Platform. We also need test users. If you like to become a test user, please let us know so we can make arrangements. This is extremely helpful for us to further develop the MULTIPLY Platform. A helpdesk function will probably depend on the final business model.

In 2019, a Summer school is scheduled which will be a continuation of the user & developer involvement. As soon as more information comes available, we will inform you. In between it would be great to stay in contact so we will think about how to arrange and exchange. Any idea you might have is appreciated.

A user forum through GitHub may be a possibility. We will publish on GitHub who is responsible for which module. On GitHub an issue tracker is up and running. This is one possibility to notify when something has been modified. Of course, since the MULTIPLY Platform is work in progress, at the moment many things don't run.

The participant list will be shared, including email addresses since nobody opposes against this. All presentations will be distributed through WeTransfer.

We recently started to share news on Twitter via @Multiply2020.

Is there any other feedback?
Ecolass would like a large research set . Is that already available such that a user can find a couple of sentinel sites and run these? Or not yet?
The planned release (in 1 – 2 months) should allow to do that. The coupling will not be available at that time yet, but different components will be.

Making it work

- How to get real time access to data
- Offer the choice to parallelize processes
- High computer power possibility to link to diverse dataset
- Guarantee 10 years of stability (to allow scientific reproducibility)
- Keep it modular and flexible
- Promote modular and local usage and development of methods
- Provide cloud computing for the retrieval of land surface parameters
- Ensure the usage of non-public but web-based usage
- Expand to other languages like R, Matlab
- Developing a business model for services

Roadmap – Processing flow MULTIPLY Platform
Since we can't improve everything at the same time, which particular aspects have priority? The answer to this, helps us to identify the roadmap on timing and documenting.

A module for pre-processing data is already an output people can work with. This is necessary for the rest and can be provided as intermediate output.

How important is it to have a final product? Or to be able to produce your own product? Do you want a final product or a piece of software that allows to produce?

It is important to have individual modules that can be used and an understanding of how to couple different modules to each other.

Within the overview of the platform some components have not been listed, such as the orchestrator (which couples the components; provides the actual implementation of the

workflow). Visualisation is more than visualisation of results. It also provides a graphical user interface and other ways/means to access the platform. These will exist in addition to what is in the flow.

If you want to incorporate a different model (e.g. 6S) you need a number of the components of the platform next to orchestrator to couple them to the inference engine.

We will continue with the development of the MULTIPLY Platform. We can also spend more time on documenting. That is what the participants requested. What do you want us to start with? What are you most interested in?

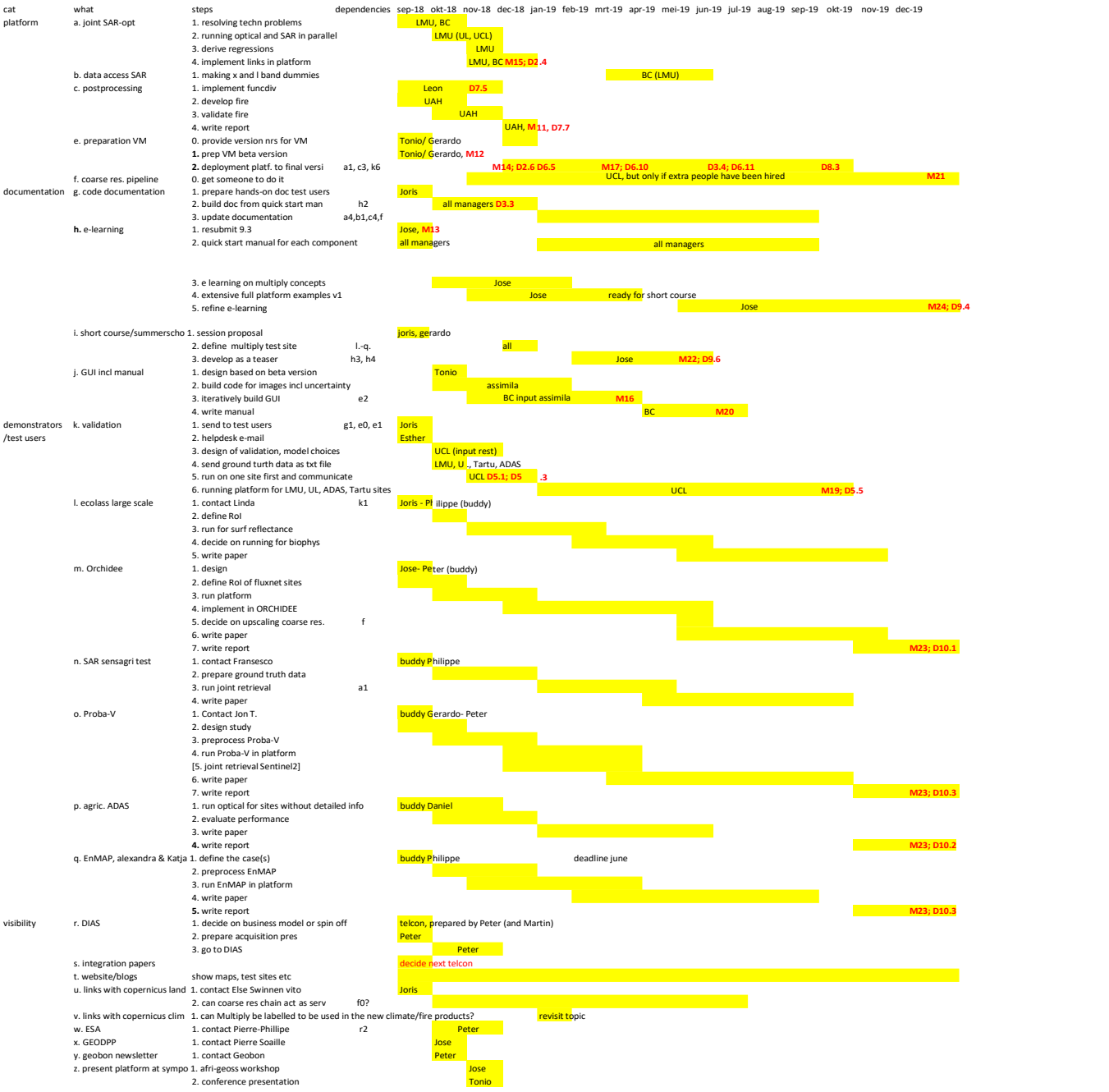
- Focus more on parameters than on the processing part.
- Ask in the first exercise: which parameters can I retrieve? You should ask users that are not experts in the field. What do you get as output? Focus on the parameters and explain them

We have delivered an initial list of models, with a description, that can be in the Platform. We did not provide a list of parameters from the point of view of the outputs, so this is a good point. We will need to improve this.

There is a parallel in parameters and validation. How far can you go? It is extremely challenging to start validation of these parameters.

We do have a number of validation sites for which we measure parameters (agricultural setting, mixed forest, pine forest). Having more validation sites is always important and better. We aim for a product to be used on a global scale. There should be coherence from a physical point of view. We might have missed important features in validation. Anything that can help us on validation is much appreciated.
Is there some sort of initiative that we can connect with to validate?

ADAS knows there is some sort of comparison run from Maryland. LMU mentions that test sites and initiatives focus either on the optical domain or the microwave domain. They do not often cover both. It would be good to have an initiative that could set up such test sites working on both domains and to provide time series instead of an individual campaign. If a spin off could be to set up campaigns to have combined test sites for validation, that would be a great outcome.
More in general: if anyone knows of a site which we could use for validation, we would appreciate those insights. It is hard to find sites to use for this kind of validation, so if you know about something, please let us know. It would help a lot. Also a review on how to properly validate various parameters and at which scale would be an added value.


cat
what
steps
dependencies
sep-18
okt-18
nov-18
dec-18
jan-19
feb-19
mrt-19
apr-19
mei-19
jun-19
jul-19
aug-19
sep-19
okt-19
nov-19
dec-19
platform
a. joint SAR-opt
1. resolving techn problems
LMU, BC
2. running optical and SAR in parallel
LMU (UL, UCL)
3. derive regressions
LMU
4. implement links in platform
LMU, BC M15; D2.4
b. data access SAR
1. making x and l band dummies
BC (LMU)
c. postprocessing
1. implement funcdiv
Leon
D7.5
2. develop fire
UAH
3. validate fire
UAH
4. write report
UAH, M11, D7.7
e. preparation VM
0. provide version nrs for VM
Tonio/ Gerardo
1. prep VM beta version
Tonio/ Gerardo, M12
2. deployment platf. to final versi
a1, c3, k6
M14; D2.6 D6.5
M17; D6.10
D3.4; D6.11
D8.3
f. coarse res. pipeline
0. get someone to do it
UCL, but only if extra people have been hired
M21
documentation
g. code documentation
1. prepare hands-on doc test users
Joris
2. build doc from quick start man
h2
all managers D3.3
3. update documentation
a4,b1,c4,f
h. e-learning
1. resubmit 9.3
Jose, M13
2. quick start manual for each component
all managers
all managers
3. e learning on multiply concepts
Jose
4. extensive full platform examples v1
Jose
ready for short course
5. refine e-learning
Jose
M24; D9.4
i. short course/summerscho
1. session proposal
joris, gerardo
2. define multiply test site
l.-q.
all
3. develop as a teaser
h3, h4
Jose
M22; D9.6
j. GUI incl manual
1. design based on beta version
Tonio
2. build code for images incl uncertainty
assimila
3. iteratively build GUI
e2
BC input assimila
M16
4. write manual
BC
M20
demonstrators /test users
k. validation
1. send to test users
g1, e0, e1
Joris
2. helpdesk e-mail
Esther
3. design of validation, model choices
UCL (input rest)
4. send ground turth data as txt file
LMU, U ., Tartu, ADAS
5. run on one site first and communicate
UCL D5.1; D5 .3
6. running platform for LMU, UL, ADAS, Tartu sites
UCL
M19; D5.5
l. ecolass large scale
1. contact Linda
k1
Joris - Ph ilippe (buddy)
2. define RoI
3. run for surf reflectance
4. decide on running for biophys
5. write paper
m. Orchidee
1. design
Jose- Peter (buddy)
2. define RoI of fluxnet sites
3. run platform
4. implement in ORCHIDEE
5. decide on upscaling coarse res.
f
6. write paper
7. write report
M23; D10.1
n. SAR sensagri test
1. contact Fransesco
buddy Philippe
2. prepare ground truth data
3. run joint retrieval
a1
4. write paper
o. Proba-V
1. Contact Jon T.
buddy Gerardo- Peter
2. design study
3. preprocess Proba-V
4. run Proba-V in platform
[5. joint retrieval Sentinel2]
6. write paper
7. write report
M23; D10.3
p. agric. ADAS
1. run optical for sites without detailed info
buddy Daniel
2. evaluate performance
3. write paper
4. write report
M23; D10.2
q. EnMAP, alexandra & Katja
1. define the case(s)
buddy Philippe
deadline june
2. preprocess EnMAP
3. run EnMAP in platform
4. write paper
5. write report
M23; D10.3
visibility
r. DIAS
1. decide on business model or spin off
telcon, prepared by Peter (and Martin)
2. prepare acquisition pres
Peter
3. go to DIAS
Peter
s. integration papers
decide next telcon
t. website/blogs
show maps, test sites etc
u. links with copernicus land
1. contact Else Swinnen vito
Joris
2. can coarse res chain act as serv
f0?
v. links with copernicus clim
1. can Multiply be labelled to be used in the new climate/fire products?
revisit topic
w. ESA
1. contact Pierre-Phillipe
r2
Peter
x. GEODPP
1. contact Pierre Soaille
Jose
y. geobon newsletter
1. contact Geobon
Peter
z. present platform at sympo
1. afri-geoss workshop
Jose
2. conference presentation
Tonio

Notes taken at the MULTIPLY Consortium Meeting at UAH on 22 and 23 January 2019

Test users:

- About 20 test users (of which 10 users got it two weeks ago).We give the test users a python notebook that runs on the VM.
- Debugging is still needed: it did not work for Lea. For Jonathan it ran, but he did not know what to do with it.
- What does it provide us? Some users will be only using parts of the platform (e.g. the pre-processing), others a simple application, others want to build their own system. We are interested in showing it provides interesting cases.
- What should it produce? shouldn't we show that in the documentation?. Now, it ends with where you can find your results.
- Buddy system: replace questionnaire by contacting buddy. Maybe even work with someone to see the language we speak and to get the interesting cases (have the test user come up with the interesting cases)
- We should also test it ourselves more extensively across the group: Nina, Daniel, Joris, Lucrecia, Leon
- Probably we need to hide more to avoid test users get confused.
- We should give recommendations on how to get to results fast (e.g. by pre-canned input data that have been atm corrected)

Afterwards get the projectmember (buddy) to acquire information

- have you been using the multiply platform

- If not, why did you not with the system?
  - Not interested
  - Not good enough
  - Too difficult
  - Not documented well enough
  - Not special enough
  - No time
  - other
- Are you still interested in having MULTIPLY provide, biophysical data for your specific application?
- Please specify what your application area (User –case studies) of interest is (drought impact monitoring, biodiversity monitoring, fire susceptibility, agricultural).
- Please indicate whether extra help is needed to being ready to analyse the MULTIPLY model for your user-case
- Acquire information regarding the user-type: *expert user, remote sensing consultant*
- Define study area and temporal extent and run the preprocessing for this (for example downloading).
- Define where the actual retrieved data is stored… (Lea)
- Ask what visualization the users would like.

After initial testing has been performed, we need extra information on:

- Are you satisfied with the configuration options of the prior engine?
- Is the procedure of changing the configuration clear?

- Regarding the upcoming deliverable 3.4 (UCL lead): which data sets would you like to include as prior data sets yourself? Which format or conventions do these data sets adhere to?
- Was the system fast enough for your purposes?
- Do you want the system locally installed? -> desktop system..

Functionality:

- SAR model still does not work with Kafka. The emulator works fine, but the parameters are not constrained. Where does this come from? We can run the model, but it is not reproducing the results. Action: Jose and Thomas together.
- SAR mainly seems to provide interesting results if combined with SAR. The solution is there, but who is going to implement that in the emulator? Thomas will take care of that and make it available for the platform. We can then make that available for interested test users (Gerardo knows one from Surrey univ).
- For forestry you would like a different RTM than SAIL.
- Fire: Tonio has defined the interface and added the code to follow the interface. It now calculates burned area (to mask the burned area and only calculate severity for the area burned) and RDR (needs still be turned into severity. That eq. was missing from the code but is available). It depends on the pre-processing but not on the vegetation par.. It goes directly from pre-processing to fire. Maybe treat them like the other par. instead of as a besides thing (hiding that we don't use the inference; reason: we don't have priors for burned areas). If you get strange results flag it, there could have been a fire. If you find there is a burned scar there, you can estimate fire severity. So, it is not really post-processing, but show it as an option within the platform by having it as an own notebook (in combi with the pre-processing); might also be nice for communication. Action: Tonio to make the notebook. How will they communicate this?
- For the post-processing still a link needs to be made to the interference. Tonio can assist Leon on this.

Validation:

- ADAS site and test case data of 2018 will be used for validation; supposed to be ready by the end of the year.
- For validation 2x forest sites and 2x agricultural sites. Speulderbos and Jarvselja: Amie and Joris. resp. UK (Assimila/ADAS) and Munich sites (by LMU). We will not use any of the sites of our catalogue (like EnMAP, ESA sites incl Barrax).
- Do the priors available in the platform and a general smoother provide a plausible estimate?
- Munich site is one of the very few sites for SAR+optical (together with the Agrisens site in Italy)
- Can we get phenotyping data (Juelich) for agriculture to improve the agricultural priors?

E-learning:

E-learning is different from documentation, as it provides the user with the information how to use the system, and get proper results from this. In documentation we only specify how particular functions should be called (sort of an API report).

The E-learning should consist of an 'example python notebook' that shows off the different prototype applications. In this regard, the e-learning platform will be developed as case-studies within the further development of the MULTIPLY Platform. In order to further facilitate knowledge uptake, the E-learning should be accompanied by a guide-through. In EGU this will be performed by an oral presenter. At other times, this could be performed by a YouTube movie, guiding people through some of these case-studies.
In this case, we require notebooks on:

- The basic concepts behind radiative transfer models
- The integration of a priori knowledge to land surface parameter retrieval. (Informative vs non-informative prior, dynamical models (how to use models)..
- Combining observations and prior information in a Bayesian retrieval scheme https://mybinder.org/v2/gh/jgomezdans/MULTIPLY_MCMC_explore/master

Possibly integrate these notebooks on the VM. Current status is half way. Highlight the need for bringing laptops.

- Create a simplified notebook.
    - Implement notebook-example with pre-canned data as an example in the e-learning environment with the same rhythm (ideally one with the same data as used for the validation: no - cannot be done soon. So that takes too long) and then indicate how you can convert it to your own case and then be allowed to continue as test user.
        - EGU2019 (biodiversity monitoring (forest -> Tartu (but might be of little use to forest experts (Joris & Lea),
        - agricultural (Assimila + ADAS)
- Perform 2 stage record notebook (2 notebooks per folder). First notebook containing basic MULTIPLY, 2nd notebook containing application.
- Make a YouTube. Action: Joris, but that is in addition to the python notebook.
- Most should go in the e-learning notebooks instead of a separate documentation. Where does e-learning stop and documentation start? Documentation should have a glossary, theoretical/conceptual application. Documentation is also classically more on the equations, while e-learning is more on how it works.
- For the three topics: 1.1 fundamental theory (complete for optical, but not for SAR), 1.2 parameter retrieval (notebook available), 1.3 how to run emulator (can move stuff from documentation to e-learning), 2.1 platform (current notebook), 2.2. pre-processing (current notebook), 2.3 priors (some examples in the Frascati notebooks and the DALEC example; using model as a prior), 3.1 cropland application (see above), forests (see above), veg traits/biodiv ind. (see LSCE; refocused for e-learning)
- buddies should keep the e-learning in mind when dealing with the test cases.
- we can track it and also publicise it to get impact.

GUI and visualisation:

- There is a difference between GUI (defining user-choices) and visualization. Two parts of the visualization. This should focus purely on biophysical parameters (it is up to the remote sensing consultant to make money on transforming these biophysical parameters to traffic light (red/green/orange) indicators.
    - Time series (with option to compare pixels against each other).
    - Uncertainty
    - Multi-parameter retrieval.

- Will we include the fire in the GUI? Or post-processing as such? GUI is just: I want to have some data. GUI is for the users that want to get into contact with python and should allow downloading the geotiffs. BC already has a GUI to work from.
- Notebooks are a starting point for those making their own code.
- The output of a notebook is a geotiff. Next version of the notebook should include basic algorithms to plot a geotiff. For some of the test cases, e.g. agriculture a dedicated visualization will be made (within or connected to a notebook).
- Visualization was linked to the GUI (and not for the notebooks). Requirements: What more than opening a geotiff; time series (one ind par. at a time, but ideally in the same plot the full ROI and multiple selections of pixels) and uncertainties (how to display that while still seeing the patterns) and a multi-dimensional retrieval. A user should also be able to select a part of the ROI for visualization and compare both parts (e.g. different fields). That would be better when using a datacube instead of geotiff. For uncertainties, a co-variance matrix showing whether you could retrieve the individual par properly (enough independent info).
- Can you do those selections in a notebook? No, not very convenient, but you can do it through a web browser (as Feng did, using standard powerful advanced Java libraries: www2.geog.ucl.ac.uk/~ucfafyi/map). With the web browser, it can query the files you need. That implies the visualization becomes separate from the GUI. Tonio would like to have them together. Solution: have the GUI start at a browser from which to call the VM and then return to the browser for visualization, querying the VM.
- The easiest solution would be to put the GUI as part of the notebook. Second easiest is to make it part of the VM (which is what Tonio envisioned and better to address the different audience). Or the other way around: the GUI is a website from where you can log in onto the VM running at the background or onto the ‘’simple sliders notebook’’ (this allows a tighter coupling to the libraries already there). Tonio prefers to start at the VM.
- Tonio will prepare a presentation for the telcon of 8 Feb on the choice to be made and the options therein on how to connect the VM, GUI and visualization.

Documentation:

- Documentation exists for:
  - SAR preprocessing
  - Atmospheric correction
  - Emulators
  - Priors

  Thomas’s to lead this activity!
- Not easy to find because of a wrong link on GitHub. Even on GitHub it is hard to find: Thomas to solve and Tonio to delete the old version to avoid wrong links. The old version has to be taken off-line and then Thomas can make a new link.
- Orchestrator not yet part of the documentation, because not used in the notebooks. But for serious applications, the orchestrator would be more useful. Maybe a wrapper is needed for the orchestrator such that the user can apply it (and documentation thereof).
- Inference engine still to be done: action by Jose.
- Coarse resolution pre-processing (is just a dataset and documentation of that is fairly simple: action Jose) and post-processing: Action for Leon.
- Documentation equals the user reference referred to in D9.3.

Coarse resolution:

- Gerardo has prepared most of the code to run the coarse resolution and someone at BC has experience in running that at large scales. So, it is feasible (although for heterogeneous areas we cannot do it and we will not solve this in this project). The only issue is the machine to run it. So, we could afford that from UCL money or convince a DIAS to host it.
- For the latter case: running coarse resolution (=modis+sentinel3+ochy) and then couple it at pre-processing to Sentinel2 would give high resolution reflectance data, which probably has a much wider user public than any other product. Maybe DIAS can run it for free or do that on JEODPP. Action Peter to follow-up.
- See five slides Jose (https://docs.google.com/presentation/d/19FF77gjp5deU686cV3XEG7R1dvc4qatPoEZ4-mmAfNQ/edit?usp=sharing) on how it can be processed. Lots of work can be done by relying on BCs software. Final product is a stored product with a global mosaic (or at least the tiles for which we want to get validation) where you can cut the bit you want. (sounds as a good solution)
- The amount of data is huge and QA4ECV was not able to do it. Therefore, you need to do it offline.

User demonstrator:

- LSCE: looking at a number of fluxnet sites (LaThuille) and retrieve data (e.g. using DALEC; the current RTM of ORCHIDEE is not yet compatible, the next version by Juliane Otto will be (in april)) as forest site with multiply-generated data. Plan is to run some tests now (e.g. compare LAI or using the multiply obs as pseudo obs) and run the real ones in April (then you have the same var., and use Multiply as pseudo obs).
- So far based on MODIS (no sentinels) which are already atm corr. There is an updated version of LaThuille till 2016, which is mostly pre-sentinels. You can run a comparison between MODIS assimilated vs sentinels assimilated data, but you cannot validate that (because of lacking flux data for that period).
- Agric.: 2017 and 2018 data. Apply it broadly to a representative set out of the 200 sites across Europe to see differences in fapar, phenology and relate that to plant performance. Can we pick up heat and drought effects of 2018?
- We can do the pre-processing for the whole tile and then do the actual processing for the field boundaries. It is important to already start the pre-processing soon to have them available and therefore to select the representative set (taking care of the tiles needed to be processed). Daniel will send the coordinates to Jon to be overlaid over the tiles and move forward from there.
- Jonathan: do the pre-processing by SIAC for proba-V. We can provide a full Multiply processing chain BRDF description for coarse resolution (for coarse res.) and then run you with a proba-V emulator or for proba-V. Another interesting comparison is comparing a proba-V retrieval with a BRDF based on proba-V or based on proba-V+sentinel3 and see how much the error bars decrease. Gerardo will take this up with Jonathan and will keep Peter in the loop on this.
- planetLabs seems more risky than proba-V as an exercise
- for enMAP we have Katja: Thomas will ask Philip on an update on this and a book chapter with Luis.

EGU session:

- Which e-learning tools do we want to present? That should focus on a use case. Maybe use the LSCE-case for the EGU-session.
- Organization: VM or Binder. Action: Jose and Joris to connect. That notebook should be tested beforehand.
- Joris will coordinate advertisement etc.: everyone should bring a laptop.
- Have a feedback form for the participants.

Traceability:

QA4ECV; Jose took care of that in the deliverable last Dec. and can be taken care in the Orchestrator. With the logging done, we can take care of the traceability.
Synthetic datasets:

- Appropriate testing with synthetic datasets has been used in testing the internals of the inference engine and the emulators (deliverable last Dec). Thanks to the large amount of data and our own ground truth data, we otherwise directly validate the entire platform.

DIAS:

- BC (Martin) created a VM for the DIAS as a way to get a platform like Multiply. Mundi and CreoDias seem promising. Action: Telcon with Mundi together with Jose. Make sure to discuss it with Martin (BC) first.
- See presentation by Tonio/Martin on how to get multiply on a DIAS system. BC will make a demonstrator for a VM to show how it can be done on a DIAS (transfer it to a DIAS to test whether it works there).
- The GUI can also be packaged in the VM.
- If you host an image and not a VM, then the costs for Multiply are much less than listed (how much? Ask Martin)
- Is google or Amazon cheaper than the DIASes?
- If we go for DIAS, we should also try to have it advertised through DIAS.
- Either go for compatibility through JEODPP and be one of the (paying) users or try to be an intermittent core in one DIAS for free (license it to a DIAS and provide capacity building services) or make a business on it. Those options may not necessarily compatible.

Outreach:

- Action: Esther to send everyone a message where to find all deliverables (the VRE?)
- Papers: the notebook in fact provides a nice basis.
- Nobody knows the ‘’scientific information service’’. Are we interested in reaching the general public? Or: how to reach the professional public? There are fairly large LinkedIn groups active and you may not need a LinkedIn page for that.
- Check the ResearchGate project site whether everyone is on it, e.g. Daniel is not on it. Send link around.
- We first need contents to communicate and distribute. You first need results, e.g. through the e-learning notebooks and you want to be sure the VM works. When it runs then have a series of communications through blogs etc..
- Put the e-learning training materials on Binder (open source; run from Github) and then they take care of it and maybe communicate that to the outside world.
- Planetarium show for Artis Zoo in Amsterdam.

- Push towards (Sub)-readit…
  - https://www.reddit.com/r/science/
  - https://www.reddit.com/r/SpaceBased/
  - https://www.reddit.com/r/remotesensing/

Papers:

- Feng has written an atm corr paper; waits for Lewis to send it off.
- We need a benchmark paper, but you also need to show some results. With Munich Sentinel2 data and possibly the UK-site, we will have enough for a paper. Jose will take the lead and then have a submittable version available at a time. That would focus on the inference engine.
- Joris will in addition take a lead on a ‘’platform-paper’’. Gerardo proposes a technical paper, why not wrap that in conceptual advances; with a limited review to start with for GMD. Also that needs to be ready by EGU.

Deliverables:

- D6.11 even in the GUI we should not hide all models, but make it easy to choose and to show the consequences thereof.
- Complication: having flexibility in choosing the forward operators also have implications for priors and for the validation of the SAR-optical coupling. Should we validate it for all combinations or indicate that it has been validated only for particular combinations. So, we should also indicate that implies that they have to use user-defined priors.
- D3.4 and D6.11 interfaces should have a common understanding to make sure they are consistent, e.g. make a library of variables indicating and defining each variable and in which forward operator they are used. Action: Jose & Tonio to stay in contact about this.

Minutes of Progress Meeting in Paris

Dates: 2 & 3 July 2019

Participants: Jon Styles, Gerardo Lopez, Nicola Pounder, Nina Raoult, Daniel Kindred, Thomas Weiss, Feng Yin, Peter van Bodegom, Joris Timmermans, Jose Gomez Dans, Grit Kirches, Thomas Ramsauer, Tonio Fincke, Philippe Peylin

1 State of Affairs

Start with deliverables and open action points.

- Delay in Joint SAR-Opt.. (this will be discussed in agenda point).
- The only thing not on the agenda concerns the data-access components for SAR. No action is required there.
- Post processing: done, no further action required.
- Preparation of VM. Will be discussed later during the meeting
- Coarse resolution pipeline: Officially we did not promise to do anything. This will be discussed in the LSE part (as coarse resolution pipeline is applied there).
- Code documentation: Priors done. SIAC done. Inference ongoing (not sure if finished by the end of August). Data access done. SAR preprocessing done. Forward operators needs to be updated. Emulators: done. Fire (Not needed as it concerns a set of regression equations). Biodiversity (needs to be updated).
- E-learning: Will be discussed in this meeting
- Summer school: This was the EGU session.
- GUI: will be discussed in the Visualization block of the meeting
- Validation: will be discussed in this meeting.
- Test use-cases: will be discussed in this meeting.
- Impact beyond the project: Guido Lemoine offered to just upload it to the JRC server: GEODPP. Possibly talk about this within these two days, specifically how to implement the platform on the DIAS systems.
- Integration Papers: Joris: not started yet, due to issues. Jose wrote a ‘data assimilation for crop growth modelling’ paper and the atmospheric correction paper has been submitted. Jose will have more time to write scientific stuff as they now have a software development team/person.
- Websites blogs; have been active
- Links with Copernicus: awaiting finalization of framework.

1.1 SAR

Prototype version (Kamka/Kaska) works (sort of). However it is not optimized yet. Within few weeks /months this version will be released, instead of the current available version (Kafka). After this UCL will in-house improve the prototype version. This was necessary because Kafka was unstable in the case of sparse observations, leading to erroneous data. Kamka/Kaska infers in a staggered (sequential) way in (similar to the Sentinel Synergy project) for particular time-windows, instead of a parallel particle filter implementation in Kafka.

For the parameter retrieval with coupled SAR/Opt assimilation, it turns out that in some cases it worked (on some fields) but not on other fields. This was not solved in the brief workshop

in London at the end of May. The optical retrievals are realistic, but when combined with SAR, it failed. Possibly, errors using Kafka were caused due to SAR model being too simplistic. There seem to be two options to continue: 1 A temporal emulator, which is basically an approximation of the Watercloud model. However it only focusses on a single parameter (LAI). But exploits the high temporal availability of the SAR data. It does not focus on soil moisture. Instead, SAR is used as an interpolator on top of a LAI retrieval. This is a possibly stable solution, e.g.to allow a remote sensing consultant to utilize the 'simplified temporal emulator'. 2. Continue with the RT modelling in which optical retrieval is done first and then (within the Inference engine) the SAR parameters of both soil moisture and LAI are retrieved. Given that this may not always work, we also need to flag simulations that do not work, using the chi square as a metric. While not everyone agreed with this option, it was decided that this option will be implemented by **Thomas W. and Jose within 5 weeks**.

The outcome may be validated at in total 5 sites; four at LMU and one at Northampton. Those can be run with measured LAI to test SAR retrieval. **Daniel** will provide coordinates of the Northampton site to Thomas such that a download of SAR data can be started and pre-processed. **Joris** will contact Philip Marzahn on comparing sentinel synergy outcomes vs. multiply outcomes for the same LMU plots.

2 User Demonstrator, presentation & discussion

2.1 Agriculture (Nicola Pounder)

Current: Kafka (discontinuous profiles using Kafka, but thoroughly debugged) vs Kaska (Variational smoother retrievals). This version also has some other add-ons like imposing a seasonal smoother. **Nicola** and **Tonio** will merge the two Kafka-versions to become a master version with all functionality and debugged performance. This will be the default version of the MULTIPLY products. If Kaska is finalized and tested, this may replace the Kafka-version, but that seems unlikely for the short time period left within the project.

2.2 MULTIPLY and ORCHIDEE (Nina Raoult)

This user demonstrator works with the coarse resolution (MODIS) MULTIPLY retrieval for LAI. This was tested in two versions of ORCHIDEE. The CMIP6-version assumes fully developed forests and has no multi-year growth embedded. Within this version, LAI is not an effective parameters. Therefore, while

LAI assimilation improves (lowering it to appropriate values), this also lowers GPP, and consequently worsened this variable.

In ORCHIDEE-CAN, it works better due to 1, a better consistency regarding LAI (because an effective LAI is modelled in exactly the same way as in MULTIPLY), and 2, a more complex modelling in Orchidee regarding LAImax.

2.3 User demonstrator to other satellites (Gerardo Lopez)

Leon Tavares (by Gerardo). Vito would like to run a LAI/FAPAR comparison with their NN approach (on the basis of Sentinel2). The user support was fine. Errors were found and fixed, but in the end, we still have errors appearing. The same errors appeared in other test user's cases. First, these errors needs to be fixed before was can continue with the test users (i.e. first tested by ourselves).Next step: remote hands-on session and installation of platform on a local VM. Such hands-on sessions are planned with multiple of our test users.

3 Testing

Peter: What is the next phase for the testing?

We will continue with the current test users, but first the errors must be fixed. Once fixed, hands-on sessions are proposed to guide the test users through the process. Possibly they may also want to test the GUI + Visualization.

Jose: MARS/JRC (and likely other users too) are not really interested in operationally using it. For this they will rely on 'outsourcing' this, but they must be able to see performance to know the potential.

Peter: Having successful users-test cases are essential to show off the potential.

4 E-learning Platform

Jose: already prepared somethings for Alcala. Nothing more has been done so far. At this moment everything is running on binder. Possibly aim to deliver it as a MOOC. For this, we could possibly ask ESA to host it as a summer school. **Jose** will finalize this in the next couple of months. Next, we can advertise the e-learning environment through reddit, geo-blogs, researchgate and others.
Validation Measurements of Leaf Pigment in Jarsjalja were presented. Results focused on differences between young/mature, peak biomass/senescent looking leaves. Although the real top leaves represent a tiny fraction (while deviating strongly from the others: less pigments during the growing season while staying green for a longer period of time), you may see them strongly in the satellite images. While abundances were not quantified, it was estimated that they might make up ¼ to 1/3 of the satellite image.
Joris: Did you measure leaf angles? No. Did you measure traits of understory? Yes: understory (mosses + shrubs), however Lea does not really trust these measurements.

For the validation report, we will focus on:

1. SIAC pre-processing validation
2. Surface reflectance for Speulderbos & Tartu, -> **Andres Kuusk** to contact Jose, **Joris** to contact Jose
3. Focus on trait-validation for LMU & Tartu, Agriculture UK
    - **Joris** to work with Tartu and Speulderbos.
    - **Jose** to work with LMU
    - **Daniel with Nicola** for the agricultural sites in the UK

Each of the validations of 3. consists (as a minimum) of PROSAIL + default priors, if time allows in addition (!) other RT models could be explored, but not on the expense of the 'default' validation. The validation will not focus on acceptance criteria, but simply focus on RMSE, and R2 to evaluate performance. DIAS + User Interface.

We have two VMs, one for DIAS with User interface and one on google engine (with EGU simplified coordinate system).

User interface is a jupyter notebook and separate from the visualisation (which is a Java Script). The User Interface is a mock-up, not yet connected to the platform.

Provide options for Roi (shape file), user Priors + indicate order of magnitude cost/time – estimate.
We don't want to run the platform as a service, but just have the image of the platform and the user can take the image to use it. Mundi is assessed directly through the open telcon cloud.
We will have to provide a manual (YOUTUBE) for users on how to make an image/VM including all the default settings that should be clicked on as it is not self-explanatory..

5 Visualisation (Nicola Pounder/ Gerardo Lopez)

The visualization is a tool that can be used locally, but also can be integrated within the VM/GUI
Gerardo: For visualization (and prior) it should be made clear if variable is transformed or not!
Jose: within the emulators, this is provided as a JSON file. Possibly also couple through variable library. We need to ensure that outputs from the VM are consistently coupled to the visualization. If the visualization becomes embedded in the Jupyter notebook, then that might be the best guarantee
Proposed improvements:

- color-scaling (make sure that 'NA values are white/transparent''), what to do with negative values. Also provide the possibility to modify the color bar
- provide different background layers (S2 RGB, S2 False color, Google maps)
- Make sure that the time-slider has only 5 (?) 'date ticks'
- Provide option to download only data within the time graph (so zoomed in, only provide those data).

6 Impact: papers, visibility, dissemination MULTIPLY Platform

If there is no platform to share, we cannot make impact (in specific writing papers!!). Get the platform running! In particular the SAR/OPT coupling! So let's get this finished within next 5 weeks.

- How to make sure we have a default VM with all the merged branches? Make sure that we only use the 'updated version Master branch' so no branched up branches. This is the case for most packages, except for the inference engine.
- For Kafka, **Tonio** and **Nicola** will investigate which versions of Kafka there are, and merge these into a single version for the VM. The focus here is not on speed, but on robustness.
- The open question is whether for SAR/OPT coupling Kaska is (technologically) needed, or if this can still be used with Kafka. It should not be a problem to take either Kafka or Kaska here (from a technical point of view) as the SAR take the output from the optical. This needs to be double checked with **Jose**.
- In August/September, teleconference to put actions to create impact (dissemination to publications).

7 Beyond Multiply

Not seen a tender for a follow-up project. There are two pathways:

- Research projects. **Daniel** mentioned that today there is a H2020 agricultural meeting. He will report back on this. Also ESA projects are an option here
- Commercial: *Philippe* Company (south of France) that sells seeds, now also needs to provide service on ‘when to sow, when to harvest’ . For this they used models that incorporate micrometeorological data. However, they are very interested in developing models that incorporate RS data. Other commercial (e.g. within Copernicus programme) options may arise.